THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. XIX, No. 3 — September, 1948

THE ANNALS OF MATHEMATICAL STATISTICS

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore 2, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves, Baltimore 2, Md., or to the Secretary of the Institute of Mathematical Statistics, P. S. Dwyer, 116 Rackham Hall, University of Michigan, Ann Arbor, Mich.

Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $8.00 inside the Western Hemisphere and $5.00 elsewhere. Single copies $3.00. Back numbers are available at $8.00 per volume or $3.00 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the act of March 3, 1879

A CLASS OF STATISTICS WITH ASYMPTOTICALLY NORMAL DISTRIBUTION[1]

By Wassily Hoeffding

Institute of Statistics, University of North Carolina

1. Summary. Let $X_1, \cdots, X_n$ be n independent random vectors, $X_\nu = (X_\nu^{(1)}, \cdots, X_\nu^{(r)})$, and $\Phi(x_1, \cdots, x_m)$ a function of $m(\leq n)$ vectors $x_\nu = (x_\nu^{(1)}, \cdots, x_\nu^{(r)})$. A statistic of the form $U = \Sigma''\Phi(X_{\alpha_1}, \cdots, X_{\alpha_m})/n(n-1) \cdots (n-m+1)$, where the sum Σ'' is extended over all permutations $(\alpha_1, \cdots, \alpha_m)$ of m different integers, $1 \leq \alpha_i \leq n$, is called a U-statistic. If $X_1, \cdots, X_n$ have the same (cumulative) distribution function (d.f.) $F(x)$, U is an unbiased estimate of the population characteristic $\theta(F) = \int \cdots \int \Phi(x_1, \cdots, x_m) dF(x_1) \cdots dF(x_m)$. $\theta(F)$ is called a regular functional of the d.f. $F(x)$. Certain optimal properties of U-statistics as unbiased estimates of regular functionals have been established by Halmos [9] (cf. Section 4).

The variance of a U-statistic as a function of the sample size n and of certain population characteristics is studied in Section 5.

It is shown that if $X_1, \cdots, X_n$ have the same distribution and $\Phi(x_1, \cdots, x_m)$ is independent of n, the d.f. of $\sqrt{n}(U - \theta)$ tends to a normal d.f. as $n \to \infty$ under the sole condition of the existence of $E\Phi^2(X_1, \cdots, X_m)$. Similar results hold for the joint distribution of several U-statistics (Theorems 7.1 and 7.2), for statistics U' which, in a certain sense, are asymptotically equivalent to U (Theorems 7.3 and 7.4), for certain functions of statistics U or U' (Theorem 7.5) and, under certain additional assumptions, for the case of the X_ν's having different distributions (Theorems 8.1 and 8.2). Results of a similar character, though under different assumptions, are contained in a recent paper by von Mises [18] (cf. Section 7).

Examples of statistics of the form U or U' are the moments, Fisher's k-statistics, Gini's mean difference, and several rank correlation statistics such as Spearman's rank correlation and the difference sign correlation (cf. Section 9). Asymptotic power functions for the non-parametric tests of independence based on these rank statistics are obtained. They show that these tests are not unbiased in the limit (Section 9f). The asymptotic distribution of the coefficient of partial difference sign correlation which has been suggested by Kendall also is obtained (Section 9h).

2. Functionals of distribution functions. Let $F(x) = F(x^{(1)}, \cdots, x^{(r)})$ be an r-variate d.f. If to any F belonging to a subset $\mathfrak{D}$ of the set of all d.f.'s in the r-dimensional Euclidean space is assigned a quantity $\theta(F)$, then $\theta(F)$ is called a

[1] Research under a contract with the Office of Naval Research for development of multivariate statistical theory.

functional of F, defined on $\mathfrak{D}$. In this paper the word functional will always mean functional of a d.f.

An infinite population may be considered as completely determined by its d.f., and any numerical characteristic of an infinite population with d.f. F that is used in statistics is a functional of F. A finite population, or sample, of size n is determined by its d.f., $S(x)$ say, and its size n. n itself is not a functional of S since two samples of different size may have the same d.f.

If $S(x^{(1)}, \cdots, x^{(r)})$ is the d.f. of a finite population, or a sample, consisting of n elements

$$x_\alpha = (x_\alpha^{(1)}, \cdots, x_\alpha^{(r)}), \qquad (\alpha = 1, \cdots, n), \tag{2.1}$$

then $nS(x^{(1)}, \cdots, x^{(r)})$ is the number of elements x_α such that

$$x_\alpha^{(1)} \leq x^{(1)}, \cdots, x_\alpha^{(r)} \leq x^{(r)}.$$

Since $S(x^{(1)}, \cdots, x^{(r)})$ is symmetric in $x_1, \cdots, x_n$, and retains its value for a sample formed from the sample (2.1) by adding one or more identical samples, the same two properties hold true for a sample functional $\theta(S)$. Most statistics in current use are functions of n and of functionals of the sample d.f.

A random sample $\{X_1, \cdots, X_n\}$ is a set of n independent random vectors

$$X_\alpha = (X_\alpha^{(1)}, \cdots, X_\alpha^{(r)}), \qquad (\alpha = 1, \cdots, n). \tag{2.2}$$

For any fixed values $x^{(1)}, \cdots, x^{(r)}$, the d.f. $S(x^{(1)}, \cdots, x^{(r)})$ of a random sample is a random variable. The functional $\theta(S)$, where S is the d.f. of the random sample, is itself a random variable, and may be called a random functional.

A remarkable application of the theory of functionals to functionals of d.f.'s has been made by von Mises [18] who considers the asymptotic distributions of certain functionals of sample d.f.'s. (Cf. also Section 7.)

3. Unbiased estimation and regular functionals. Consider a functional $\theta = \theta(F)$ of the r-variate d.f. $F(x) = F(x^{(1)}, \cdots, x^{(r)})$, and suppose that for some sample size n, θ admits an unbiased estimate for any d.f. F in $\mathfrak{D}$. That is, if $X_1, \cdots, X_n$ are n independent random vectors with the same d.f. F, there exists a function $\varphi(x_1, \cdots, x_n)$ of n vector arguments (2.1) such that the expected value of $\varphi(X_1, \cdots, X_n)$ is equal to $\theta(F)$, or

$$\int \cdots \int \varphi(x_1, \cdots, x_n)\, dF(x_1) \cdots dF(x_n) = \theta(F) \tag{3.1}$$

for every F in $\mathfrak{D}$. Here and in the sequel, when no integration limits are indicated, the integral is extended over the entire space of $x_1, \cdots, x_n$. The integral is understood in the sense of Stieltjes-Lebesgue.

The estimate $\varphi(x_1, \cdots, x_n)$ of $\theta(F)$ is called unbiased over $\mathfrak{D}$.

A functional $\theta(F)$ of the form (3.1) will be referred to as *regular over* $\mathfrak{D}$.[2]

---

[2] This is an adaptation to functionals of d.f.'s of the term "regular functional" used by Volterra [21].

Thus, the functionals regular over $\mathfrak{D}$ are those admitting an unbiased estimate over $\mathfrak{D}$.

If $\theta(F)$ is regular over $\mathfrak{D}$, let $m(\leq n)$ be the smallest sample size for which there exists an unbiased estimate $\Phi(x_1, \cdots, x_m)$ of θ over $\mathfrak{D}$:

$$\theta(F) = \int \cdots \int \Phi(x_1, \cdots, x_m)\, dF(x_1) \cdots dF(x_m) \tag{3.2}$$

for any F in $\mathfrak{D}$. Then m will be called *the degree over* $\mathfrak{D}$ of the regular functional $\theta(F)$.

If the expected value of $\varphi(X_1, \cdots, X_n)$ is equal to $\theta(F)$ whenever it exists, $\varphi(x_1, \cdots, x_n)$ will be called a *distribution-free unbiased estimate* (d-f. u.e.) of $\theta(F)$. The degree of $\theta(F)$ over the set $\mathfrak{D}_0$ of d.f.'s F for which the right hand side of (3.1) exists will be simply termed the *degree* of $\theta(F)$.

A regular functional of degree 1 over $\mathfrak{D}$ is called a linear regular functional over $\mathfrak{D}$. If $\theta(F)$ has the same value for all F in $\mathfrak{D}$, $\theta(F)$ may be termed a regular functional of degree zero over $\mathfrak{D}$.

Any function $\Phi(x_1, \cdots, x_m)$ satisfying (3.2) will be referred to as a *kernel* of the regular functional $\theta(F)$.

For any regular functional $\theta(F)$ there exists a kernel $\Phi_0(x_1, \cdots, x_m)$ symmetric in $x_1, \cdots, x_m$. For if $\Phi(x_1, \cdots, x_m)$ is a kernel of $\theta(F)$,

$$\Phi_0(x_1, \cdots, x_m) = \frac{1}{m!} \Sigma \Phi(x_{\alpha_1}, \cdots, x_{\alpha_m}), \tag{3.3}$$

where the sum is taken over all permutations $(\alpha_1, \cdots, \alpha_m)$ of $(1, \cdots, m)$, is a symmetric kernel of $\theta(F)$.

If $\theta_1(F)$ and $\theta_2(F)$ are two regular functionals of degrees m_1 and m_2 over $\mathfrak{D}$, then the sum $\theta_1(F) + \theta_2(F)$ and the product $\theta_1(F)\theta_2(F)$ are regular functionals of degrees $\leq m = \text{Max}\,(m_1, m_2)$ and $\leq m_1 + m_2$, respectively, over $\mathfrak{D}$. For if $\Phi_i(x_1, \cdots, x_{m_i})$ is a kernel of $\theta_i(F)$, $(i = 1, 2)$, then

$$\theta_1(F) + \theta_2(F) = \int \cdots \int \{\Phi_1(x_1, \cdots, x_{m_1}) + \Phi_2(x_1, \cdots, x_{m_2})\} \, dF(x_1) \cdots dF(x_m)$$

and

$$\theta_1(F)\theta_2(F) = \int \cdots \int \Phi_1(x_1, \cdots, x_{m_1})\Phi_2(x_{m_1+1}, \cdots, x_{m_1+m_2}) \, dF(x_1) \cdots dF(x_{m_1+m_2}).$$

More generally, *a polynomial in regular functionals is itself a regular functional.* Examples of linear regular functionals are the moments about the origin,

$$\mu'_{\nu_1, \cdots, \nu_r} = \int \cdots \int (x^{(1)})^{\nu_1} \cdots (x^{(r)})^{\nu_r} \, dF(x^{(1)}, \cdots, x^{(r)}).$$

A moment about the mean is a polynomial in moments μ' about 0, and hence a regular functional over the set $\mathfrak{D}_0$ of d.f.'s for which it exists (cf. Halmos [9]). For instance, the variance of $X^{(1)}$,

$$\sigma^2 = \int\int ((x_1^{(1)})^2 - x_1^{(1)} x_2^{(1)})\, dF(x_1^{(1)})\, dF(x_2^{(1)})$$

is a regular functional of degree 2. A symmetrical kernel of σ^2 is $(x^{(1)} - x^{(2)})^2/2$. If $\mathfrak{D}$ is the set of univariate d.f.'s with mean μ and existing second moment, σ^2 is a linear regular functional of F over $\mathfrak{D}$, since then we have

$$\sigma^2 = \int (x_1^{(1)} - \mu)^2\, dF(x_1^{(1)}).$$

The function

$$v = \frac{1}{n(n-1)} \sum_{\alpha \neq \beta} \frac{1}{2} (x_\alpha^{(1)} - x_\beta^{(1)})^2 = \frac{1}{n-1} \sum_\alpha \left(x_\alpha^{(1)} - \frac{1}{n} \sum_\beta x_\beta^{(1)} \right)^2$$

is a distribution-free unbiased estimate of σ^2. The function

$$\Gamma\left(\frac{n-1}{2}\right) \sqrt{\frac{n-1}{2}}\, \sqrt{v} / \Gamma\left(\frac{n}{2}\right)$$

is known to be an unbiased estimate of σ over the set of univariate normal d.f.'s, but it is not a d.-f. u.e.

4. U-statistics. Let $x_1, \cdots, x_n$ be a sample of n vectors (2.1) and $\Phi(x_1, \cdots, x_m)$ a function of $m (\leq n)$ vector arguments. Consider the function of the sample,

$$U = U(x_1, \cdots, x_n) = \frac{1}{n(n-1) \cdots (n-m+1)} \Sigma'' \Phi(x_{\alpha_1}, \cdots, x_{\alpha_m}), \tag{4.1}$$

where Σ'' stands for summation over all permutations $(\alpha_1, \cdots, \alpha_m)$ of m integers such that

$$1 \leq \alpha_i \leq n, \qquad \alpha_i \neq \alpha_j \text{ if } i \neq j, \qquad (i, j = 1, \cdots, m). \tag{4.2}$$

U is the average of the values of Φ in the set of ordered subsets of m members of the sample (2.1). U is symmetric in $x_1, \cdots, x_n$.

Any statistic of the form (4.1) will be called a *U-statistic*. Any function $\Phi(x_1, \cdots, x_m)$ satisfying (4.1) will be referred to as a *kernel* of the statistic U.

If $\Phi(x_1, \cdots, x_m)$ is a kernel of a regular functional $\theta(F)$ defined on a set $\mathfrak{D}$, then U is an unbiased estimate of $\theta(F)$ over $\mathfrak{D}$:

$$\theta(F) = \int \cdots \int U(x_1, \cdots, x_n)\, dF(x_1) \cdots dF(x_n) \tag{4.3}$$

for every F in $\mathfrak{D}$.

For $n = m$, U reduces to the symmetric kernel (3.3) of $\theta(F)$.

From a recent paper by Halmos [9] it follows for the case of univariate d.f.'s $(r = 1)$:

If $\theta(F)$ is a regular functional of degree m over a set $\mathfrak{D}$ containing all purely discontinuous d.f.'s, U is the only unbiased estimate over $\mathfrak{D}$ which is symmetric in $x_1, \cdots, x_n$, and U has the least variance among all unbiased estimates over $\mathfrak{D}$.

These results and the proofs given by Halmos can easily be extended to the multivariate case $(r > 1)$.

Combining (3.3) and (4.1) we may write a U-statistic in the form

$$U(x_1, \cdots, x_n) = \binom{n}{m}^{-1} \Sigma' \Phi_0(x_{\alpha_1}, \cdots, x_{\alpha_m}), \tag{4.4}$$

where the kernel Φ_0 is symmetric in its m vector arguments and the sum Σ' is extended over all subscripts α such that

$$1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_m \leq n.$$

Another statistic frequently used for estimating $\theta(F)$ is $\theta(S)$, where $S = S(x)$ is the d.f. of the sample (2.1). If S is substituted for F in (3.2), we have

$$\theta(S) = \frac{1}{n^m} \sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_m=1}^{n} \Phi(x_{\alpha_1}, \cdots, x_{\alpha_m}). \tag{4.5}$$

In particular, the sample moments have this form; their kernel Φ is obtained by the method described in section 3.

If $m = 1$, $\theta(S) = U$. If $m = 2$,

$$\theta(S) = \frac{n-1}{n} U + \frac{1}{n} \left\{ \frac{1}{n} \sum_{\alpha=1}^{n} \Phi(x_\alpha, x_\alpha) \right\},$$

and $\theta(S)$ is a linear function of U-statistics with coefficients depending on n. This is easily seen to be true for any m. In general $\theta(S)$ is not an unbiased estimate of $\theta(F)$. If, however, the expected value of $\theta(S)$ exists for every F in $\mathfrak{D}$, we have

$$E\{\theta(S)\} = \theta(F) + O(n^{-1}),$$

and the estimate $\theta(S)$ of $\theta(F)$ may be termed unbiased in the limit over $\mathfrak{D}$.

Numerous statistics in current use have the form of, or can be expressed in terms of U-statistics. From what was said above about moments as regular functionals, it is easy to obtain U-statistics which are d.-f. u.e.'s of the moments about the mean of any order (cf. Halmos [9]). Fisher's k-statistics are U-statistics, as follows from their definition as unbiased estimates of the cumulants, symmetric in the sample values. Another example is Gini's mean difference

$$\frac{1}{n(n-1)} \sum_{\alpha \neq \beta} | x_\alpha^{(1)} - x_\beta^{(1)} |.$$

More examples, in particular of rank correlation statistics, will be given in section 9.

5. The variance of a U-statistic. Let $X_1, \cdots, X_n$ be n independent random vectors with the same d.f. $F(x) = F(x^{(1)}, \cdots, x^{(r)})$, and let

$$U = U(X_1, \cdots, X_n) = \binom{n}{m}^{-1} \Sigma' \Phi(X_{\alpha_1}, \cdots, X_{\alpha_m}), \tag{5.1}$$

where $\Phi(x_1, \cdots, x_m)$ is symmetric in $x_1, \cdots, x_m$ and Σ' has the same meaning as in (4.4). Suppose that the function Φ does not involve n.

If $\theta = \theta(F)$ is defined by (3.2), we have

$$E\{U\} = E\{\Phi(X_1, \cdots, X_m)\} = \theta.$$

Let

$$\Phi_c(x_1, \cdots, x_c) = E\{\Phi(x_1, \cdots, x_c, X_{c+1}, \cdots, X_m)\}, \quad (c = 1, \cdots, m), \tag{5.2}$$

where $x_1, \cdots, x_c$ are arbitrary fixed vectors and the expected value is taken with respect to the random vectors $X_{c+1}, \cdots, X_m$. Then

$$\Phi_{c-1}(x_1, \cdots, x_{c-1}) = E\{\Phi_c(x_1, \cdots, x_{c-1}, X_c)\}, \tag{5.3}$$

and

$$E\{\Phi_c(X_1, \cdots, X_c)\} = \theta, \qquad (c = 1, \cdots, m). \tag{5.4}$$

Define

$$\Psi(x_1, \cdots, x_m) = \Phi(x_1, \cdots, x_m) - \theta, \tag{5.5}$$

$$\Psi_c(x_1, \cdots, x_c) = \Phi_c(x_1, \cdots, x_c) - \theta, \qquad (c = 1, \cdots, m). \tag{5.6}$$

We have

$$\Psi_{c-1}(x_1, \cdots, x_{c-1}) = E\{\Psi_c(x_1, \cdots, x_{c-1}, X_c)\}, \tag{5.7}$$

$$E\{\Psi_c(X_1, \cdots, X_c)\} = E\{\Psi(X_1, \cdots, X_m)\} = 0, \qquad (c = 1, \cdots, m). \tag{5.8}$$

Suppose that the variance of $\Psi_c(X_1, \cdots, X_c)$ exists, and let

$$\zeta_0 = 0, \qquad \zeta_c = E\{\Psi_c^2(X_1, \cdots, X_c)\}, \qquad (c = 1, \cdots, m). \tag{5.9}$$

We have

$$\zeta_c = E\{\Phi_c^2(X_1, \cdots, X_c)\} - \theta^2. \tag{5.10}$$

$\zeta_c = \zeta_c(F)$ is a polynomial in regular functionals of F, and hence itself a regular functional of F (of degree $\leq 2m$).

If, for some parent distribution $F = F_0$ and some integer d, we have $\zeta_d(F_0) = 0$, this means that $\Psi_d(X_1, \cdots, X_d) = 0$ with probability 1. By (5.7) and (5.9), $\zeta_d = 0$ implies $\zeta_1 = \cdots = \zeta_{d-1} = 0$.

If $\zeta_1(F_0) = 0$, we shall say that the regular functional $\theta(F)$ is *stationary*[3] for $F = F_0$. If

(5.11) $$\zeta_1(F_0) = \cdots = \zeta_d(F_0) = 0, \qquad \zeta_{d+1}(F_0) > 0, \qquad (1 \leq d \leq m),$$

$\theta(F)$ will be called *stationary of order* d for $F = F_0$.

If $(\alpha_1, \cdots, \alpha_m)$ and $(\beta_1, \cdots, \beta_m)$ are two sets of m different integers, $1 \leq \alpha_i$, $\beta_i \leq n$, and c is the number of integers common to the two sets, we have, by the symmetry of Ψ,

(5.12) $$E\{\Psi(X_{\alpha_1}, \cdots, X_{\alpha_m})\Psi(X_{\beta_1}, \cdots, X_{\beta_m})\} = \zeta_c .$$

If the variance of U exists, it is equal to

$$\sigma^2(U) = \binom{n}{m}^{-2} E\{\Sigma'\Psi(X_{\alpha_1}, \cdots, X_{\alpha_m})\}^2$$
$$= \binom{n}{m}^{-2} \sum_{c=0}^{m} \Sigma^{(c)} E\{\Psi(X_{\alpha_1}, \cdots, X_{\alpha_m})\Psi(X_{\beta_1}, \cdots, X_{\beta_m})\},$$

where $\Sigma^{(c)}$ stands for summation over all subscripts such that

$$1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_m \leq n, \qquad 1 \leq \beta_1 < \beta_2 < \cdots < \beta_m \leq n,$$

and exactly c equations

$$\alpha_i = \beta_j$$

are satisfied. By (5.12), each term in $\Sigma^{(c)}$ is equal to ζ_c. The number of terms in $\Sigma^{(c)}$ is easily seen to be

$$\frac{n(n-1)\cdots(n-2m+c+1)}{c!(m-c)!(m-c)!} = \binom{m}{c}\binom{n-m}{m-c}\binom{n}{m},$$

and hence, since $\zeta_0 = 0$,

(5.13) $$\sigma^2(U) = \binom{n}{m}^{-1} \sum_{c=1}^{m} \binom{m}{c}\binom{n-m}{m-c}\zeta_c .$$

When the distributions of $X_1, \cdots, X_n$ are different, $F_\nu(x)$ being the d.f. of X_ν, let

(5.14) $$\theta_{\alpha_1, \cdots, \alpha_m} = E\{\Phi(X_{\alpha_1}, \cdots, X_{\alpha_m})\},$$

$$\Psi_{c(\alpha_1, \cdots, \alpha_c)\beta_1, \cdots, \beta_{m-c}}(x_1, \cdots, x_c)$$

(5.15) $$= E\{\Phi(x_1, \cdots, x_c, X_{\beta_1} \ldots, X_{\beta_{m-c}})\} - \theta_{\alpha_1, \cdots, \alpha_c, \beta_1, \cdots, \beta_{m-c}},$$

$$(c = 1, \cdots, m),$$

[3] According to the definition of the derivative of a functional (cf. Volterra [21]; for functionals of d.f.'s cf. von Mises [18]), the function $m(m-1)\ldots(m-d+1)\,\Psi_d(x_1 \ldots x_d)$, which is a functional of F, is a d-th derivative of $\theta(F)$ with respect to F at the "point" F of the space of d.f.'s.

$$\zeta_{c(\alpha_1,\cdots,\alpha_c)\beta_1,\cdots,\beta_{m-c};\gamma_1,\cdots,\gamma_{m-c}}$$

$$= E\{\Psi_{c(\alpha_1,\cdots,\alpha_c)\beta_1,\cdots,\beta_{m-c}}(X_{\alpha_1}, \cdots, X_{\alpha_c})\Psi_{c(\alpha_1,\cdots,\alpha_c)\gamma_1,\cdots,\gamma_{m-c}}(X_{\alpha_1}, \cdots, X_{\alpha_c})\} \tag{5.16}$$

$$\zeta_{c,n} = \frac{c!(m-c)!(m-c)!}{n(n-1)\cdots(n-2m+c+1)} \Sigma\zeta_{c(\alpha_1,\cdots,\alpha_c)\beta_1,\cdots,\beta_{m-c};\gamma_1,\cdots,\gamma_{m-c}} \tag{5.17}$$

where the sum is extended over all subscripts α, β, γ such that

$$1 \le \alpha_1 < \cdots < \alpha_c \le n, \quad 1 \le \beta_1 < \cdots < \beta_{m-c} \le n, \quad 1 \le \gamma_1 < \cdots \gamma_{m-c} \le n,$$

$$\alpha_i \ne \beta_j, \qquad \alpha_i \ne \gamma_j, \qquad \beta_i \ne \gamma_j.$$

Then the variance of U is equal to

$$\sigma^2(U) = \binom{n}{m}^{-1} \sum_{c=1}^{m} \binom{m}{c}\binom{n-m}{m-c} \zeta_{c,n}. \tag{5.18}$$

Returning to the case of identically distributed X's, we shall now prove some inequalities satisfied by $\zeta_1, \cdots, \zeta_m$ and $\sigma^2(U)$ which are contained in the following theorems:

THEOREM 5.1 *The quantities $\zeta_1, \cdots, \zeta_m$ as defined by* (5.9) *satisfy the inequalities*

$$0 \le \frac{\zeta_c}{c} \le \frac{\zeta_d}{d} \qquad \textit{if } 1 \le c < d \le m. \tag{5.19}$$

THEOREM 5.2 *The variance $\sigma^2(U_n)$ of a U-statistic $U_n = U(X_1, \cdots, X_n)$, where $X_1, \cdots, X_n$ are independent and identically distributed, satisfies the inequalities*

$$\frac{m^2}{n}\zeta_1 \le \sigma^2(U_n) \le \frac{m}{n}\zeta_m. \tag{5.20}$$

$n\sigma^2(U_n)$ is a decreasing function of n,

$$(n+1)\sigma^2(U_{n+1}) \le n\sigma^2(U_n), \tag{5.21}$$

which takes on its upper bound $m\zeta_m$ for $n = m$ and tends to its lower bound $m^2\zeta_1$ as n increases:

$$\sigma^2(U_m) = \zeta_m, \tag{5.22}$$

$$\lim_{n\to\infty} n\sigma^2(U_n) = m^2\zeta_1. \tag{5.23}$$

If $E\{U_n\} = \theta(F)$ is stationary of order $\ge d - 1$ for the d.f. of X_α, (5.20) *may be replaced by*

$$\frac{m}{d} K_n(m, d)\zeta_d \le \sigma^2(U_n) \le K_n(m, d)\zeta_m, \tag{5.24}$$

where

$$K_n(m, d) = \binom{n}{m}^{-1} \sum_{c=d}^{m} \binom{m-1}{c-1}\binom{n-m}{m-c}. \tag{5.25}$$

We postpone the proofs of Theorems 5.1 and 5.2.

(5.13) and (5.19) imply that a necessary and sufficient condition for the existence of $\sigma^2(U)$ is the existence of

$$\zeta_m = E\{\Phi^2(X_1, \cdots, X_m)\} - \theta^2 \tag{5.26}$$

or that of $E\{\Phi^2(X_1, \cdots, X_m)\}$.

If $\zeta_1 > 0$, $\sigma^2(U)$ is of order n^{-1}.

If $\theta(F)$ is stationary of order d for $F = F_0$, that is, if (5.11) is satisfied, $\sigma^2(U)$ is of order n^{-d-1}. Only if, for some $F = F_0$, $\theta(F)$ is stationary of order m, where m is the degree of $\theta(F)$, we have $\sigma^2(U) = 0$, and U is equal to a constant with probability 1.

For instance, if $\theta(F_0) = 0$, the functional $\theta^2(F)$ is stationary for $F = F_0$. Other examples of stationary "points" of a functional will be found in section 9d.

For proving Theorem 5.1 we shall require the following:

LEMMA 5.1. *If*

$$\delta_d = \zeta_d - \binom{d}{1}\zeta_{d-1} + \binom{d}{2}\zeta_{d-2} \cdots + (-1)^{d-1}\binom{d}{d-1}\zeta_1, \tag{5.27}$$

we have

$$\delta_d \geq 0, \qquad (d = 1, \cdots, m) \tag{5.28}$$

and

$$\zeta_d = \delta_d + \binom{d}{1}\delta_{d-1} + \cdots + \binom{d}{d-1}\delta_1. \tag{5.29}$$

PROOF. (5.29) follows from (5.27) by induction.

For proving (5.28) let

$$\eta_0 = \theta^2, \qquad \eta_c = E\{\Phi_c^2(X_1, \cdots, X_c)\}, \qquad (c = 1, \cdots, m).$$

Then, by (5.10),

$$\zeta_c = \eta_c - \eta_0,$$

and on substituting this in (5.27) we have

$$\delta_d = \sum_{c=0}^{d} (-1)^{d-c}\binom{d}{c}\eta_c.$$

From (5.9) it is seen that (5.28) is true for $d = 1$. Suppose that (5.28) holds for $1, \cdots, d-1$. Then (5.28) will be shown to hold for d.

Let

$$\bar{\Phi}_0(x_1) = \Phi_1(x_1) - \theta, \qquad \bar{\Phi}_c(x_1, x_2, \cdots, x_{c+1})$$

$$= \Phi_{c+1}(x_1, \cdots, x_{c+1}) - \Phi_c(x_2, \cdots, x_{c+1}), \qquad (c = 1, \cdots, d-1).$$

For an arbitrary fixed x_1, let

$$\bar{\eta}_c(x_1) = E\{\bar{\Phi}_c^2(x_1, X_2, \cdots, X_{c+1})\}, \qquad (c = 0, \cdots, d-1).$$

Then, by induction hypothesis,

$$\bar{\delta}_{d-1}(x_1) = \sum_{c=0}^{d-1} (-1)^{d-1-c} \binom{d-1}{c} \bar{\eta}_c(x_1) \geq 0$$

for any fixed x_1.

Now,

$$E\{\bar{\eta}_c(X_1)\} = \eta_{c+1} - \eta_c,$$

and hence

$$E\{\bar{\delta}_{d-1}(X_1)\} = \sum_{c=0}^{d-1} (-1)^{d-1-c} \binom{d-1}{c} (\eta_{c+1} - \eta_c) = \sum_{c=0}^{d} (-1)^{d-c} \binom{d}{c} \eta_c = \delta_d.$$

The proof of Lemma 5.1 is complete.

Proof of Theorem 5.1. By (5.29) we have for $c < d$

$$c\zeta_d - d\zeta_c = c \sum_{a=1}^{d} \binom{d}{a} \delta_a - d \sum_{a=1}^{c} \binom{c}{a} \delta_a$$

$$= \sum_{a=1}^{c} \left[c\binom{d}{a} - d\binom{c}{a} \right] \delta_a + c \sum_{a=c+1}^{d} \binom{d}{a} \delta_a. \tag{5.30}$$

From (5.28), and since $c\binom{d}{a} - d\binom{c}{a} \geq 0$ if $1 \leq a \leq c \leq d$, it follows that each term in the two sums of (5.30) is not negative. This, in connection with (5.9) proves Theorem 5.1.

Proof of Theorem 5.2. From (5.19) we have

$$c\zeta_1 \leq \zeta_c \leq \frac{c}{m} \zeta_m, \qquad (c = 1, \cdots, m).$$

Applying these inequalities to each term in (5.13) and using the identity

$$\binom{n}{m}^{-1} \sum_{c=1}^{m} c \binom{m}{c} \binom{n-m}{m-c} = \frac{m^2}{n}, \tag{5.31}$$

we obtain (5.20).

(5.22) and (5.23) follow immediately from (5.13).

For (5.21) we may write

$$D_n \geq 0, \tag{5.32}$$

where

$$D_n = n\sigma^2(U_n) - (n + 1)\sigma^2(U_{n+1}).$$

Let

$$D_n = \sum_{c=1}^{m} d_{n,c}\zeta_c.$$

Then we have from (5.13)

$$d_{n,c} = n\binom{m}{c}\binom{n-m}{m-c}\binom{n}{m}^{-1} - (n+1)\binom{m}{c}\binom{n+1-m}{m-c}\binom{n+1}{m}^{-1}, \tag{5.33}$$

or

$$d_{n,c} = \binom{m}{c}\binom{n-m+1}{m-c}(n-m+1)^{-1}\binom{n}{m}^{-1}\{(c-1)n - (m-1)^2\},$$
$$(1 \leq c \leq m \leq n).$$

Putting

$$c_0 = 1 + \left[\frac{(m-1)^2}{n}\right],$$

where $[u]$ denotes the largest integer $\leq u$, we have

$$d_{n,c} \leq 0 \qquad \text{if } c \leq c_0,$$
$$d_{n,c} > 0 \qquad \text{if } c > c_0.$$

Hence, by (5.19),

$$d_{n,c}\zeta_c \geq \frac{1}{c_0}\zeta_{c_0} c d_{n,c}, \qquad (c = 1, \cdots, m),$$

and

$$D_n \geq \frac{1}{c_0}\zeta_{c_0}\sum_{c=1}^{m} c d_{n,c}.$$

By (5.33) and (5.31), the latter sum vanishes. This proves (5.32).

For the stationary case $\zeta_1 = \cdots = \zeta_{d-1} = 0$, (5.24) is a direct consequence of (5.13) and (5.19). The proof of Theorem 5.2 is complete.

6. The covariance of two U-statistics. Consider a set of g U-statistics,

$$U^{(\gamma)} = \binom{n}{m(\gamma)}^{-1}\Sigma'\Phi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}}), \qquad (\gamma = 1, \cdots, g),$$

each $U^{(\gamma)}$ being a function of the same n independent, identically distributed random vectors $X_1, \cdots, X_n$. The function $\Phi^{(\gamma)}$ is assumed to be symmetric in its $m(\gamma)$ arguments $(\gamma = 1, \cdots, g)$.

Let

$$E\{U^{(\gamma)}\} = E\{\Phi^{(\gamma)}(X_1, \cdots, X_{m(\gamma)})\} = \theta^{(\gamma)}, \qquad (\gamma = 1, \cdots, g);$$

$$\Psi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)}) = \Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)}) - \theta^{(\gamma)}, \qquad (\gamma = 1, \cdots, g); \tag{6.1}$$

$$\Psi_c^{(\gamma)}(x_1, \cdots, x_c) = E\{\Psi^{(\gamma)}(x_1, \cdots, x_c, X_{c+1}, \cdots, X_{m(\gamma)})\}, \qquad (c = 1, \cdots, m(\gamma); \gamma = 1, \cdots, g); \tag{6.2}$$

$$\zeta_c^{(\gamma,\delta)} = E\{\Psi_c^{(\gamma)}(X_1, \cdots, X_c)\Psi_c^{(\delta)}(X_1, \cdots, X_c)\}, \qquad (\gamma, \delta = 1, \cdots, g). \tag{6.3}$$

If, in particular, $\gamma = \delta$, we shall write

$$\zeta_c^{(\gamma)} = \zeta_c^{(\gamma,\gamma)} = E\{\Psi_c^{(\gamma)}(X_1, \cdots, X_c)\}^2. \tag{6.4}$$

Let

$$\sigma(U^{(\gamma)}, U^{(\delta)}) = E\{(U^{(\gamma)} - \theta^{(\gamma)})(U^{(\delta)} - \theta^{(\delta)})\}$$

be the covariance of $U^{(\gamma)}$ and $U^{(\delta)}$.

In a similar way as for the variance, we find, if $m(\gamma) \leq m(\delta)$,

$$\sigma(U^{(\gamma)}, U^{(\delta)}) = \binom{n}{m(\gamma)}^{-1} \sum_{c=1}^{m(\gamma)} \binom{m(\delta)}{c}\binom{n - m(\delta)}{m(\gamma) - c} \zeta_c^{(\gamma,\delta)}. \tag{6.5}$$

The right hand side is easily seen to be symmetric in γ, δ.

For $\gamma = \delta$, (6.5) is the variance of $U^{(\gamma)}$ (cf. (5.13)).

We have from (5.23) and (6.5)

$$\lim_{n\to\infty} n\sigma^2(U^{(\gamma)}) = m^2(\gamma)\zeta_1^{(\gamma)},$$

$$\lim_{n\to\infty} n\sigma(U^{(\gamma)}, U^{(\delta)}) = m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)}.$$

Hence, if $\zeta_1^{(\gamma)} \neq 0$ and $\zeta_1^{(\delta)} \neq 0$, the product moment correlation $\rho(U^{(\gamma)}, U^{(\delta)})$ between $U^{(\gamma)}$ and $U^{(\delta)}$ tends to the limit

$$\lim_{n\to\infty} \rho(U^{(\gamma)}, U^{(\delta)}) = \frac{\zeta_1^{(\gamma,\delta)}}{\sqrt{\zeta_1^{(\gamma)}\zeta_1^{(\delta)}}}. \tag{6.6}$$

7. Limit theorems for the case of identically distributed X_α's. We shall now study the asymptotic distribution of U-statistics and certain related functions. In this section the vectors X_α will be assumed to be identically distributed. An extension to the case of different parent distributions will be given in section 8.

Following Cramér [2, p. 83] we shall say that a sequence of d.f.'s $F_1(x)$, $F_2(x)$, $\cdots$ converges to a d.f. $F(x)$ if $\lim F_n(x) = F(x)$ in every point at which the one-dimensional marginal limiting d.f.'s are continuous.

Let us recall (cf. Cramér [2, p. 312]) that a g-variate normal distribution is called non-singular if the rank r of its covariance matrix is equal to g, and singular if $r < g$.

The following lemma will be used in the proofs.

LEMMA 7.1. *Let $V_1, V_2, \cdots$ be an infinite sequence of random vectors $V_n = (V_n^{(1)}, \cdots, V_n^{(g)})$, and suppose that the d.f. $F_n(v)$ of V_n tends to a d.f. $F(v)$ as $n \to \infty$. Let $V_n^{(\gamma)\prime} = V_n^{(\gamma)} + d_n^{(\gamma)}$, where*

$$\lim_{n\to\infty} E\{d_n^{(\gamma)}\}^2 = 0, \qquad (\gamma = 1, \cdots, g). \tag{7.1}$$

Then the d.f. of $V_n' = (V_n^{(1)\prime}, \cdots, V_n^{(g)\prime})$ tends to $F(v)$.

This is an immediate consequence of the well-known fact that the d.f. of V_n' tends to $F(v)$ if $d_n^{(\gamma)}$ converges in probability to 0 (cf. Cramér [2, p. 299]), since the fulfillment of (7.1) is sufficient for the latter condition.

THEOREM 7.1. *Let $X_1, \cdots, X_n$ be n independent, identically distributed random vectors,*

$$X_\alpha = (X_\alpha^{(1)}, \cdots, X_\alpha^{(r)}), \qquad (\alpha = 1, \cdots, n).$$

Let

$$\Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)}), \qquad (\gamma = 1, \cdots, g),$$

be g real-valued functions not involving n, $\Phi^{(\gamma)}$ being symmetric in its $m(\gamma)$ $(\leq n)$ vector arguments $x_\alpha = (x_\alpha^{(1)}, \cdots, x_\alpha^{(r)})$, $(\alpha = 1, \cdots, m(\gamma);\ \gamma = 1, \cdots, g)$. Define

$$U^{(\gamma)} = \binom{n}{m(\gamma)}^{-1} \sum{}' \Phi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}}), \qquad (\gamma = 1, \cdots, g), \tag{7.2}$$

where the summation is over all subscripts such that $1 \leq \alpha_1 < \cdots < \alpha_{m(\gamma)} \leq n$. Then, if the expected values

$$\theta^{(\gamma)} = E\{\Phi^{(\gamma)}(X_1, \cdots, X_{m(\gamma)})\}, \qquad (\gamma = 1, \cdots, g), \tag{7.3}$$

and

$$E\{\Phi^{(\gamma)}(X_1, \cdots, X_{m(\gamma)})\}^2, \qquad (\gamma = 1, \cdots, g), \tag{7.4}$$

exist, the joint d.f. of

$$\sqrt{n}(U^{(1)} - \theta^{(1)}), \cdots, \sqrt{n}(U^{(g)} - \theta^{(g)})$$

tends, as $n \to \infty$, to the g-variate normal d.f. with zero means and covariance matrix $(m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)})$, where $\zeta_1^{(\gamma,\delta)}$ is defined by (6.3). The limiting distribution is non-singular if the determinant $|\zeta_1^{(\gamma,\delta)}|$ is positive.

Before proving Theorem 7.1, a few words may be said about its meaning and its relation to well-known results.

For $g = 1$, Theorem 7.1 states that the distribution of a U-statistic tends, under certain conditions, to the normal form. For $m = 1$, U is the sum of n inde-

pendent random variables, and in this case Theorem 7.1 reduces to the Central Limit Theorem for such sums. For $m > 1$, U is a sum of random variables which, in general, are not independent. Under certain assumptions about the function $\Phi(x_1, \cdots, x_m)$ the asymptotic normality of U can be inferred from the Central Limit Theorem by well-known methods. If, for instance, Φ is a polynomial (as in the case of the k-statistics or the unbiased estimates of moments), U can be expressed as a polynomial in moments about the origin which are sums of independent random variables, and for this case the tendency to normality of U can easily be shown (cf. Cramér [2, p. 365]).

Theorem 7.1 generalizes these results, stating that in the case of independent and identically distributed X_α's the existence of $E\{\Phi^2(X_1, \cdots, X_m)\}$ is sufficient for the asymptotic normality of U. No regularity conditions are imposed on the function Φ. This point is important for some applications (cf. section 9).

Theorem 7.1 and the following theorems of sections 7 and 8 are closely related to recent results of von Mises [18] which were published after this paper was essentially completed. It will be seen below (Theorem 7.4) that the limiting distribution of $\sqrt{n}[U - \theta(F)]$ is the same as that of $\sqrt{n}[\theta(S) - \theta(F)]$ (cf. (4.5)) if the variance of $\theta(S)$ exists. $\theta(S)$ is a differentiable statistical function in the sense of von Mises, and by Theorem I of [18], $\sqrt{n}[\theta(S) - \theta(F)]$ is asymptotically normal if certain conditions are satisfied. It will be found that in certain cases, for instance if the kernel Φ of θ is a polynomial, the conditions of the theorems of sections 7 and 8 are somewhat weaker than those of von Mises' theorem. Though von Mises' paper is concerned with functionals of univariate d.f.'s only, its results can easily be extended to the multivariate case.

For the particular case of a discrete population (where F is a step function), U and $\theta(S)$ are polynomials in the sample frequencies, and their asymptotic distribution may be inferred from the fact that the joint distribution of the frequencies tends to the normal form (cf. also von Mises [18]).

In Theorem 7.1 the functions $\Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)})$ are supposed to be symmetric. Since, as has been seen in section 4, any U-statistic with non-symmetric kernel can be written in the form (4.4) with a symmetric kernel, this restriction is not essential and has been made only for the sake of convenience. Moreover, in the condition of the existence of $E\{\Phi^2(X_1, \cdots, X_m)\}$, the symmetric kernel may be replaced by a non-symmetric one. For, if Φ is non-symmetric, and Φ_0 is the symmetric kernel defined by (3.3), $E\{\Phi_0^2(X_1, \cdots, X_m)\}$ is a linear combination of terms of the form $E\{\Phi(X_{\alpha_1}, \cdots, X_{\alpha_m})\, \Phi\, (X_{\beta_1}, \cdots, X_{\beta_m})\}$, whose existence follows from that of $E\{\Phi^2(X_1, \cdots, X_m)\}$ by Schwarz's inequality.

If the regular functional $\theta(F)$ is stationary for $F = F_0$, that is, if $\zeta_1 = \zeta_1(F_0) = 0$ (cf. section 5), the limiting normal distribution of $\sqrt{n}(U - \theta)$ is, according to Theorem 7.1, singular, that is, its variance is zero. As has been seen in section 5, $\sigma^2(U)$ need not be zero in this case, but may be of some order n^{-c}, $(c = 2, 3, \cdots, m)$, and the distribution of $n^{c/2}(U - \theta)$ may tend to a limiting form which is not normal. According to von Mises [18], it is a limiting distribution of type c, $(c = 2, 3, \cdots)$.

According to Theorem 5.2, $\sigma^2(U)$ exceeds its asymptotic value $m^2\zeta_1/n$ for any finite n. Hence, if we apply Theorem 7.1 for approximating the distribution of U when n is large but finite, we underestimate the variance of U. For many applications this is undesirable, and for such cases the following theorem, which is an immediate consequence of Theorem 7.1, will be more useful.

THEOREM 7.2. *Under the conditions of Theorem* 7.1, *and if*

$$\zeta_1^{(\gamma)} > 0, \qquad (\gamma = 1, \cdots, g),$$

the joint d.f. of

$$(U^{(1)} - \theta^{(1)})/\sigma(U^{(1)}), \cdots, (U^{(g)} - \theta^{(g)})/\sigma(U^{(g)})$$

tends, as $n \to \infty$, to the g-variate normal d.f. with zero means and covariance matrix $(\rho^{(\gamma,\delta)})$, *where*

$$\rho^{(\gamma,\delta)} = \lim_{n\to\infty} \frac{\sigma(U^{(\gamma)}, U^{(\delta)})}{\sigma(U^{(\gamma)})\sigma(U^{(\delta)})} = \frac{\zeta_1^{(\gamma,\delta)}}{\sqrt{\zeta_1^{(\gamma)}\zeta_1^{(\delta)}}}, \qquad (\gamma, \delta = 1, \cdots, g).$$

PROOF OF THEOREM 7.1. The existence of (7.4) entails that of

$$\zeta_m^{(\gamma)} = E\{\Phi^{(\gamma)}(X_1, \cdots, X_{m(\gamma)})\}^2 - (\theta^{(\gamma)})^2$$

which, by (5.19), (5.20) and (6.6), is sufficient for the existence of

$$\zeta_1^{(\gamma)}, \cdots, \zeta_{m-1}^{(\gamma)}, \quad \text{of} \quad \sigma^2(U^{(\gamma)}), \quad \text{and of} \quad \zeta_1^{(\gamma,\delta)} \leq \sqrt{\zeta_1^{(\gamma)}\zeta_1^{(\delta)}}.$$

Now, consider the g quantities

$$Y^{(\gamma)} = \frac{m(\gamma)}{\sqrt{n}} \sum_{\alpha=1}^{n} \Psi_1^{(\gamma)}(X_\alpha), \qquad (\gamma = 1, \cdots, g)$$

where $\Psi_1^{(\gamma)}(x)$ is defined by (6.2). $Y^{(1)}, \cdots, Y^{(g)}$ are sums of n independent, random variables with zero means, whose covariance matrix, by virtue of (6.3), is

$$\{\sigma(Y^{(\gamma)}, Y^{(\delta)})\} = \{m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)}\}. \tag{7.5}$$

By the Central Limit Theorem for vectors (cf. Cramér [1, p. 112]), the joint d.f. of $(Y^{(1)}, \cdots, Y^{(g)})$ tends to the normal g-variate d.f. with the same means and covariances.

Theorem 7.1 will be proved by showing that the g random variables

$$Z^{(\gamma)} = \sqrt{n}(U^{(\gamma)} - \theta^{(\gamma)}), \qquad (\gamma = 1, \cdots, g), \tag{7.6}$$

have the same joint limiting distribution as $Y^{(1)}, \cdots, Y^{(g)}$.

According to Lemma 7.1 it is sufficient to show that

$$\lim_{n\to\infty} E(Z^{(\gamma)} - Y^{(\gamma)})^2 = 0, \qquad (\gamma = 1, \cdots, n). \tag{7.7}$$

For proving (7.7), write

$$E\{Z^{(\gamma)} - Y^{(\gamma)}\}^2 = E\{Z^{(\gamma)}\}^2 + E\{Y^{(\gamma)}\}^2 - 2E\{Z^{(\gamma)}Y^{(\gamma)}\}. \tag{7.8}$$

By (5.13) we have

$$E\{Z^{(\gamma)}\}^2 = n\sigma^2(U^{(\gamma)}) = m^2(\gamma)\zeta_1^{(\gamma)} + O(n^{-1}), \tag{7.9}$$

and from (7.5),

$$E\{Y^{(\gamma)}\}^2 = m^2(\gamma)\zeta_1^{(\gamma)}. \tag{7.10}$$

By (7.2) and (6.1) we may write for (7.6)

$$Z^{(\gamma)} = \sqrt{n}\binom{n}{m(\gamma)}^{-1} \Sigma' \, \Psi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}}),$$

and hence

$$E\{Z^{(\gamma)}Y^{(\gamma)}\} = m(\gamma)\binom{n}{m(\gamma)}^{-1} \sum_{\alpha=1}^{n} \sum{}' E\{\Psi_1^{(\gamma)}(X_\alpha)\Psi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}})\}.$$

The term

$$E\{\Psi_1^{(\gamma)}(X_\alpha)\Psi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}})\}$$

is $= \zeta_1^{(\gamma)}$ if

$$\alpha_1 = \alpha \quad \text{or} \quad \alpha_2 = \alpha \cdots \quad \text{or} \quad \alpha_{m(\gamma)} = \alpha \tag{7.11}$$

and 0 otherwise. For a fixed α, the number of sets $\{\alpha_1, \cdots, \alpha_{m(\gamma)}\}$ such that $1 \le \alpha_1 < \cdots < \alpha_{m(\gamma)} \le n$ and (7.11) is satisfied, is $\binom{n-1}{m(\gamma)-1}$. Thus,

$$E\{Z^{(\gamma)}Y^{(\gamma)}\} = m(\gamma)\binom{n}{m(\gamma)}^{-1} n \binom{n-1}{m(\gamma)-1}\zeta_1^{(\gamma)} = m^2(\gamma)\zeta_1^{(\gamma)}. \tag{7.12}$$

On inserting (7.9), (7.10), and (7.12) in (7.8), we see that (7.7) is true.

The concluding remark in Theorem 7.1 is a direct consequence of the definition of a non-singular distribution. The proof of Theorem 7.1 is complete.

Theorems 7.1 and 7.2 deal with the asymptotic distribution of $U^{(1)}, \cdots, U^{(g)}$, which are unbiased estimates of $\theta^{(1)}, \cdots, \theta^{(g)}$. The unbiasedness of a statistic is, of course, irrelevant for its asymptotic behavior, and the application of Lemma 7.1 leads immediately to the following extension of Theorem 7.1 to a larger class of statistics.

THEOREM 7.3. *Let*

$$U^{(g)\prime} = U^{(g)} + \frac{b_n^{(\gamma)}}{\sqrt{n}}, \qquad (\gamma = 1, \cdots, g), \tag{7.13}$$

where $U^{(\gamma)}$ is defined by (7.2) and $b_n^{(\gamma)}$ is a random variable. If the conditions of Theorem 7.1 are satisfied, and $\lim E\{b_n^{(\gamma)}\}^2 = 0$, $(\gamma = 1, \cdots, g)$, *then the joint distribution of*

$$\sqrt{n}(U^{(1)\prime} - \theta^{(1)}), \cdots, \sqrt{n}(U^{(g)\prime} - \theta^{(g)})$$

tends to the normal distribution with zero means and covariance matrix

$$\{m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)}\}.$$

This theorem applies, in particular, to the regular functionals $\theta(S)$ of the sample d.f.,

$$\theta(S) = \frac{1}{n^m} \sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_m=1}^{n} \Phi(X_{\alpha_1}, \cdots, X_{\alpha_m}),$$

in the case that the variance of $\theta(S)$ exists. For we may write

$$n^m\theta(S) = \binom{n}{m} U + \Sigma^* \Phi(X_{\alpha_1}, \cdots, X_{\alpha_m}),$$

where the sum Σ^* is extended over all m-tuplets $(\alpha_1, \cdots, \alpha_m)$ in which at least one equality $\alpha i = \alpha_j (i \neq j)$ is satisfied. The number of terms in Σ^* is of order n^{m-1}. Hence

$$\theta(S) - U = \frac{1}{n} D,$$

where the expected value $E\{D^2\}$, whose existence follows from that of $\sigma^2\{\theta(S)\}$, is bounded for $n \to \infty$. Thus, if we put $U^{(\gamma)'} = \theta^{(\gamma)}(S)$, the conditions of Theorem 7.3 are fulfilled. We may summarize this result as follows:

THEOREM 7.4. *Let $X_1, \cdots, X_n$ be a random sample from an r-variate population with d.f. $F(x) = F(x^{(1)}, \cdots, x^{(r)})$, and let*

$$\theta^{(\gamma)}(F) = \int \cdots \int \Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)})\, dF(x_1) \cdots dF(x_{m^{(\gamma)}}), \qquad (\gamma = 1, \cdots, g),$$

be g regular functionals of F, where $\Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)})$ is symmetric in the vectors $x_1, \cdots, x_{m(\gamma)}$ and does not involve n. If $S(x)$ is the d.f. of the random sample, and if the variance of

$$\theta^{(\gamma)}(S) = \frac{1}{n^m} \sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_{m(\gamma)}=1}^{n} \Phi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}})$$

exists, the joint d.f. of

$$\sqrt{n}\{\theta^{(1)}(S) - \theta^{(1)}(F)\}, \cdots, \sqrt{n}\{\theta^{(g)}(S) - \theta^{(g)}(F)\}$$

tends to the g-variate normal d.f. with zero means and covariance matrix

$$\{m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)}\}.$$

The following theorem is concerned with the asymptotic distribution of a function of statistics of the form U or U'.

THEOREM 7.5. *Let $(U') = (U^{(1)'}, \cdots, U^{(g)'})$ be a random vector, where $U^{(\gamma)'}$ is defined by (7.13), and suppose that the conditions of Theorem 7.3 are satisfied. If the function $h(y) = h(y^{(1)}, \cdots, y^{(g)})$ does not involve n and is continuous together with its second order partial derivatives in some neighborhood of the point $(y) = (\theta) = (\theta^{(1)}, \cdots, \theta^{(g)})$, then the distribution of the random variable $\sqrt{n}\{h(U') - h(\theta)\}$ tends to the normal distribution with mean zero and variance*

$$\sum_{\gamma=1}^{g} \sum_{\delta=1}^{g} m(\gamma)m(\delta) \left(\frac{\partial h(y)}{\partial y^{(\gamma)}}\right)_{y=\theta} \left(\frac{\partial h(y)}{\partial y^{(\delta)}}\right)_{y=\theta} \zeta_1^{(\gamma,\delta)}.$$

Theorem 7.5 follows from Theorem 7.3 in exactly the same way as the theorem on the asymptotic distribution of a function of moments follows from the fact of their asymptotic normality; cf. Cramér [2, p. 366]. We shall therefore omit the proof of Theorem 7.5. Since any moment whose variance exists has the form $U' = \theta(S)$ (cf. section 4 and Theorem 7.4), Theorem 7.5 is a generalization of the theorem on a function of moments.

8. Limit theorems for $U(X_1, \cdots, X_n)$ when the X_α's have different distributions. The limit theorems of the preceding section can be extended to the case when the X_α's have different distributions. We shall only prove an extension to this case of Theorem 7.1 (or 7.2), confining ourselves, for the sake of simplicity, to the distribution of a single U-statistic.

The extension of Theorems 7.3 and 7.5 with $g = 1$ to this case is immediate. One has only to replace the reference to Theorem 7.1 by that to the following Theorem 8.1, and θ and ζ_1 by $E\{U\}$ and $\zeta_{1,n}$.

THEOREM 8.1. *Let $X_1, \cdots, X_n$ be n independent random vectors of r components, X_α having the d.f. $F_\alpha(x) = F_\alpha(x^{(1)}, \cdots, x^{(r)})$. Let $\Phi(x_1, \cdots, x_m)$ be a function symmetric in its m vector arguments $x_\beta = (x_\beta^{(1)}, \cdots, x_\beta^{(r)})$ which does not involve n, and let*

$$\bar{\Psi}_{1(\nu)}(x) = \binom{n-1}{m-1}^{-1} \sum_{(\neq \nu)}{}' \Psi_{1(\nu)\alpha_1,\cdots,\alpha_{m-1}}(x), \qquad (\nu = 1, \cdots, n), \tag{8.1}$$

where Ψ is defined by (5.15), *and the summation is extended over all subscripts α such that*

$$1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_{m-1} \leq n, \qquad \alpha_i \neq \nu, \qquad (i = 1, \cdots, m).$$

Suppose that there is a number A such that for every $n = 1, 2, \cdots$

$$\int \cdots \int \Phi^2(x_1, \cdots, x_m)\, dF_{\alpha_1}(x_1) \cdots dF_{\alpha_m}(x_m) < A, \tag{8.2}$$

$$(1 \leq \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_m \leq n),$$

that

$$E\,|\,\bar{\Psi}^3_{1(\nu)}(X_\nu)\,| < \infty, \qquad (\nu = 1, 2, \cdots, n), \tag{8.3}$$

and

$$\lim_{n\to\infty} \sum_{\nu=1}^{n} E\,|\,\bar{\Psi}^3_{1(\nu)}(X_\nu)\,| \Big/ \left\{\sum_{\nu=1}^{n} E\{\bar{\Psi}^2_{1(\nu)}(X_\nu)\}\right\}^{3/2} = 0. \tag{8.4}$$

Then, as $n \to \infty$, the d.f. of $(U - E\{U\})/\sigma(U)$ tends to the normal d.f. with mean 0 *and variance* 1.

The proof is similar to that of Theorem 7.1.

Let

$$W = \frac{m}{n} \sum_{\nu=1}^{n} \bar{\Psi}_{1(\nu)}(X_\nu).$$

It will be shown that
(a) the d.f. of

$$V = \frac{W - E\{W\}}{\sigma(W)}$$

tends to the normal d.f. with mean 0 and variance 1, and that
(b) the d.f. of

$$V' = \frac{U - E\{U\}}{\sigma(U)}$$

tends to the same limit as the d.f. of V.

Part (a) follows immediately from (8.3) and (8.4) by Liapounoff's form of the Central Limit Theorem.

According to Lemma 7.1, (b) will be proved when it is shown that

$$\lim_{n\to\infty} E\{V' - V\}^2 = \lim\left\{2 - 2\,\frac{\sigma(U, W)}{\sigma(U)\sigma(W)}\right\} = 0$$

or

$$\lim_{n\to\infty} \frac{\sigma(U, W)}{\sigma(U)\sigma(W)} = 1. \tag{8.5}$$

Let c be an integer, $1 \leq c \leq m$, and write

$$\boldsymbol{x} = (x_1, \cdots, x_c), \qquad \boldsymbol{y} = (y_1, \cdots, y_{m-c}), \qquad \boldsymbol{z} = (z_1, \cdots, z_{m-c})$$

$$F_{(\alpha)}(\boldsymbol{x}) = F_{\alpha_1}(x_1) \cdots F_{\alpha_c}(x_c), \qquad F_{(\beta)}(\boldsymbol{y}) = F_{\beta_1}(y_1) \cdots F_{\beta_{m-c}}(y_{m-c}),$$

$$F_{(\gamma)}(\boldsymbol{z}) = F_{\gamma_1}(z_1) \cdots F_{\gamma_{m-c}}(z_{m-c}).$$

Then, by Schwarz's inequality,

$$\int \cdots \int \Phi(\boldsymbol{x}, \boldsymbol{y})\Phi(\boldsymbol{x}, \boldsymbol{z})\, dF_{(\alpha)}(\boldsymbol{x})\, dF_{(\beta)}(\boldsymbol{y})\, dF_{(\gamma)}(\boldsymbol{z})$$

$$\leq \left\{\int \cdots \int \Phi^2(\boldsymbol{x}, \boldsymbol{y})\, dF_{(\alpha)}(\boldsymbol{x})\, dF_{(\beta)}(\boldsymbol{y}) \cdot \int \cdots \int \Phi^2(\boldsymbol{x}, \boldsymbol{z})\, dF_{(\alpha)}(\boldsymbol{x})\, dF_{(\gamma)}(\boldsymbol{z})\right\}^{\frac{1}{2}},$$

which, by (8.2), is $< A$ for any set of subscripts.

By the inequality for moments, $\theta_{\alpha_1,\cdots,\alpha_m}$, as defined by (5.14), is also uniformly bounded, and applying these inequalities to (5.16), it follows that there exists a number B such that

$$|\zeta_{c(\alpha_1,\cdots,\alpha_c)\beta_1,\cdots,\beta_{m-c};\,\gamma_1,\cdots,\gamma_{m-c}}| < B, \qquad (c = 1, \cdots, m), \tag{8.6}$$

for every set of subscripts satisfying the inequalities

$$\alpha_g \neq \alpha_h, \quad \beta_g \neq \beta_h, \quad \gamma_g \neq \gamma_h \quad \text{if } g \neq h, \quad \alpha_i \neq \beta_j, \quad \alpha_i \neq \gamma_j,$$

$$(i = 1, \cdots, c;\ j = 1, \cdots, m - c).$$

Now, we have

$$E\{W\} = 0$$

and

$$\sigma^2(W) = \frac{m^2}{n^2} \sum_{\nu=1}^{n} E\{\bar{\Psi}^2_{1(\nu)}(X_\nu)\} \tag{8.7}$$

or, inserting (8.1) and recalling (5.16),

$$\sigma^2(W) = \frac{m^2}{n^2} \binom{n-1}{m-1}^{-2} \sum_{\nu=1}^{n} \sum_{(\neq \nu)}' \sum_{(\neq \nu)}' \zeta_{1(\nu)\alpha_1, \cdots, \alpha_{m-1}; \beta_1, \cdots, \beta_{m-1}}, \tag{8.8}$$

the two sums Σ' being over $\alpha_1 < \cdots < \alpha_{m-1}$, $(\alpha_i \neq \nu)$, and $\beta_1 < \cdots < \beta_{m-1}$, $(\beta_i \neq \nu)$, respectively. By (5.17), the sum of the terms whose subscripts $\nu, \alpha_1, \cdots, \alpha_{m-1}, \beta_1, \cdots, \beta_{m-1}$ are all different is equal to

$$\frac{n(n-1)\cdots(n-2m+2)}{(m-1)!(m-1)!} \zeta_{1,n} = n \binom{n-1}{m-1}\binom{n-m}{m-1} \zeta_{1,n}.$$

The number of the remaining terms is of order n^{2m-2}. Since, by (8.6), they are uniformly bounded, we have

$$\sigma^2(W) = \frac{m^2}{n} \zeta_{1,n} + O(n^{-2}). \tag{8.9}$$

Similarly, we have from (5.18)

$$\sigma^2(U) = \frac{m^2}{n} \zeta_{1,n} + O(n^{-2}),$$

and hence

$$\sigma(U) = \sigma(W) + O(n^{-1}). \tag{8.10}$$

The covariance of U and W is

$$\sigma(U, W) = \binom{n}{m}^{-1} \frac{m}{n} \sum_{\nu=1}^{n} \sum{}' E\{\bar{\Psi}_{1(\nu)}(X_\nu)\Psi_{m(\alpha_1, \cdots, \alpha_m)}(X_{\alpha_1}, \cdots, X_{\alpha_m})\}. \tag{8.11}$$

All terms except those in which one of the α's $= \nu$, vanish, and for the remaining ones we have, for fixed $\alpha_1, \cdots, \alpha_m$,

$$\begin{aligned} E\{\bar{\Psi}_{1(\nu)}(X_\nu)\Psi_{m(\alpha_1, \cdots, \alpha_m)}(X_{\alpha_1}, \cdots, X_{\alpha_m})\} \\ &= \binom{n-1}{m-1}^{-1} \sum_{(\neq \nu)}' E\{\Psi_{1(\nu)\beta_1, \cdots, \beta_{m-1}}(X_\nu)\Psi_{1(\nu)\gamma_1, \cdots, \gamma_{m-1}}(X_\nu)\} \\ &= \binom{n-1}{m-1}^{-1} \sum_{(\neq \nu)}' \zeta_{1(\nu)\beta_1, \cdots, \beta_{m-1}; \gamma_1, \cdots, \gamma_{m-1}} \end{aligned}$$

where the summation sign refers to the β's, and $\gamma_1, \cdots, \gamma_{m-1}$ are the α's that are $\neq \nu$. Inserting this in (8.11) and comparing the result with (8.8), we see that

$$\sigma(U, W) = \sigma^2(W). \tag{8.12}$$

From (8.12) and (8.10) we have

$$\frac{\sigma(U, W)}{\sigma(U)\sigma(W)} = \frac{\sigma(W)}{\sigma(U)} = \frac{n\sigma(W)}{n\sigma(W) + O(1)}.$$

Comparing condition (8.4) with (8.7), we see that we must have $n\sigma(W) \to \infty$ as $n \to \infty$. This shows the truth of (8.5). The proof of Theorem 8.1 is complete.

For some purposes the following corollary of Theorem 8.1 will be useful, where the conditions (8.2), (8.3), and (8.4) are replaced by other conditions which are more restrictive, but easier to apply.

THEOREM 8.2. *Theorem* 8.1 *holds if the conditions* (8.2), (8.3), *and* (8.4) *are replaced by the following:*
There exist two positive numbers C, D *such that*

$$(8.13) \qquad \int \cdots \int |\Phi^3(x_1, \cdots, x_m)| \, dF_{\alpha_1}(x_1) \cdots dF_{\alpha_m}(x_m) < C$$

for $\alpha_i = 1, 2, \cdots, (i = 1, \cdots, m)$, *and*

$$(8.14) \qquad \zeta_{1(\nu)\alpha_1, \cdots, \alpha_{m-1}; \beta_1, \cdots, \beta_{m-1}} > D$$

for any subscripts satisfying

$$1 \le \alpha_1 < \alpha_2 < \cdots < \alpha_{m-1}, \qquad 1 \le \beta_1 < \beta_2 < \cdots < \beta_{m-1}, \qquad 1 \le \nu \ne \alpha_i, \beta_i.$$

We have to show that (8.2), (8.3), and (8.4) follow from (8.13) and (8.14).

(8.13) implies (8.2) by the inequality for moments. By a reasoning analogous to that used in the previous proof, applying Hölder's inequality instead of Schwarz's inequality, it follows from (8.13) that

$$(8.15) \qquad E\,|\bar{\Psi}^3_{1(\nu)}(X_\nu)| < C'.$$

On the other hand, by (8.7), (8.8), and (8.14),

$$(8.16) \qquad \sum_{\nu=1}^{n} E\{\bar{\Psi}^2_{1(\nu)}(X_\nu)\} > nD.$$

(8.15) and (8.16) are sufficient for the fulfillment of (8.4).

9. Applications to particular statistics.

(a) *Moments and functions of moments.* It has been seen in section 4 that the k-statistics and the unbiased estimates of moments are U-statistics, while the sample moments are regular functionals of the sample d.f. By Theorems 7.1, 8.1, and 7.4 these statistics are asymptotically normally distributed, and by Theorem 7.5 the same is true for a function of moments, if the respective conditions are satisfied. These results are not new (cf., for example, Cramér [2]).

(b) *Mean difference and coefficient of concentration.* If $Y_1, \cdots, Y_n$ are n independent real-valued random variables, Gini's mean difference (without repetition) is defined by

$$d = \frac{1}{n(n-1)} \sum_{\alpha \ne \beta} |Y_\alpha - Y_\beta|.$$

If the Y_α's have the same distribution F, the mean of d is

$$\delta = \int\int |y_1 - y_2| \, dF(y_1) \, dF(y_2),$$

and the variance, by (5.13) is

$$\sigma^2(d) = \frac{2}{n(n-1)} \{2\zeta_1(\delta)(n-2) + \zeta_2(\delta)\},$$

where

$$\zeta_1(\delta) = \int \left\{ \int |y_1 - y_2| \, dF(y_2) \right\}^2 dF(y_1) - \delta^2, \tag{9.1}$$

$$\zeta_2(\delta) = \int\int (y_1 - y_2)^2 \, dF(y_1) \, dF(y_2) - \delta^2 = 2\sigma^2(Y) - \delta^2. \tag{9.2}$$

The notation $\zeta_1(\delta)$, $\zeta_2(\delta)$ serves to indicate the relation of these functionals of F to the functional $\delta(F)$; δ is here merely the symbol of the functional, not a particular value of it. In a similar way we shall write $\Phi(y_1, y_2 \mid \delta) \doteq |y_1 - y_2|$, etc. When there is danger of confusing $\zeta_1(\delta)$ with $\zeta_1(F)$, we may write $\zeta_1(F \mid \delta)$.

U. S. Nair [19] has evaluated $\sigma^2(d)$ for several particular distributions.

By Theorem 7.1, $\sqrt{n}(d - \delta)$ is asymptotically normal if $\zeta_2(\delta)$ exists.

If $Y_1, \cdots, Y_n$ do not assume negative values, the coefficient of concentration (cf. Gini [8]) is defined by

$$G = \frac{d}{2\bar{Y}},$$

where $\bar{Y} = \Sigma Y_\alpha / n$. G is a function of two U-statistics. If the Y_α's are identically distributed, if $E\{Y^2\}$ exists, and if $\mu = E\{Y\} > 0$, then, by Theorem 7.5, $\sqrt{n}(G - \delta/2\mu)$ tends to be normally distributed with mean 0 and variance

$$\frac{\delta^2}{4\mu^4} \zeta_1(\mu) - \frac{\delta}{\mu^3} \zeta_1(\mu, \delta) + \frac{1}{\mu^2} \zeta_1(\delta),$$

where

$$\zeta_1(\mu) = \int y^2 \, dF(y) - \mu^2 = \sigma^2(Y),$$

$$\zeta_1(\mu, \delta) = \int\int y_1 |y_1 - y_2| \, dF(y_1) \, dF(y_2) - \mu\delta,$$

and $\zeta_1(\delta)$ is given by (9.1).

(c) *Functions of ranks and of the signs of variate differences.* Let $s(u)$ be the signum function,

$$s(u) = \begin{cases} -1 \text{ if } u < 0; \\ 0 \text{ if } u = 0; \\ 1 \text{ if } u > 0, \end{cases} \tag{9.3}$$

and let

$$c(u) = \tfrac{1}{2}\{1 + s(u)\} = \begin{cases} 0 \text{ if } u < 0; \\ \tfrac{1}{2} \text{ if } u = 0; \\ 1 \text{ if } u > 0. \end{cases} \tag{9.4}$$

If

$$x_\alpha = (x_\alpha^{(1)}, \cdots, x_\alpha^{(r)}), \qquad (\alpha = 1, \cdots, n)$$

is a sample of n vectors of r components, we may define the *rank* $R_\alpha^{(i)}$ of $x_\alpha^{(i)}$ by

$$\begin{aligned} R_\alpha^{(i)} &= \tfrac{1}{2} + \sum_{\beta=1}^{n} c(x_\alpha^{(i)} - x_\beta^{(i)}) \\ &= \frac{n+1}{2} + \tfrac{1}{2} \sum_{\beta=1}^{n} s(x_\alpha^{(i)} - x_\beta^{(i)}), \qquad (i = 1, \cdots, r). \end{aligned} \tag{9.5}$$

If the numbers $x_1^{(i)}, x_2^{(i)}, \cdots, x_n^{(i)}$ are all different, the smallest of them has rank 1, the next smallest rank 2, etc. If some of them are equal, the rank as defined by (9.5) is known as the mid-rank.

Any function of the ranks is a function of expressions $c(x_\alpha^{(i)} - x_\beta^{(i)})$ or $s(x_\alpha^{(i)} - x_\beta^{(i)})$.

Conversely, since

$$s(x_\alpha^{(i)} - x_\beta^{(i)}) = s(R_\alpha^{(i)} - R_\beta^{(i)}),$$

any function of expressions $s(x_\alpha^{(i)} - x_\beta^{(i)})$ or $c(x_\alpha^{(i)} - x_\beta^{(i)})$ is a function of the ranks.

Consider a regular functional $\theta(F)$ whose kernel $\Phi(x_1, \cdots, x_m)$ depends only on the signs of the variate differences,

$$s(x_\alpha^{(i)} - x_\beta^{(i)}), \qquad (\alpha, \beta = 1, \cdots, m; i = 1, \cdots, r). \tag{9.6}$$

The corresponding U-statistic is a function of the ranks of the sample variates.

The function Φ can take only a finite number of values, $c_1, \cdots, c_N$, say. If $\pi_i = P\{\Phi = c_i\}$, $(i = 1, \cdots, N)$, we have

$$\theta = c_1\pi_1 + \cdots + c_N\pi_N, \qquad \sum_{i=1}^{N} \pi_i = 1.$$

π_i is a regular functional whose kernel $\Phi_i(x_1, \cdots, x_m)$ is equal to 1 or 0 according to whether $\Phi = c_i$ or $\neq c_i$. We have

$$\Phi = c_1\Phi_1 + \cdots + c_N\Phi_N .$$

In order that $\theta(F)$ exist, the c_i must be finite, and hence Φ is bounded. Therefore, $E\{\Phi^2\}$ exists, and if $X_1, X_2, \cdots$ are identically distributed, the d.f. of $\sqrt{n}(U - \theta)$ tends, by Theorem 7.1, to a normal d.f. which is non-singular if $\zeta_1 > 0$.

In the following we shall consider several examples of such functionals.

(d) *Difference sign correlation.* Consider the bivariate sample

$$(x_1^{(1)}, x_1^{(2)}), (x_2^{(1)}, x_2^{(2)}), \cdots, (x_n^{(1)}, x_n^{(2)}). \tag{9.7}$$

To each two members of this sample corresponds a pair of signs of the differences of the respective variables,

$$s(x_\alpha^{(1)} - x_\beta^{(1)}), \; s(x_\alpha^{(2)} - x_\beta^{(2)}), \qquad (\alpha \neq \beta; \alpha, \beta = 1, \cdots, n). \tag{9.8}$$

(9.8) is a population of $n(n-1)$ pairs of difference signs. Since

$$\sum_{\alpha \neq \beta} s(x_\alpha^{(i)} - x_\beta^{(i)}) = 0, \qquad (i = 1, 2),$$

the covariance t of the difference signs (9.8) is

$$t = \frac{1}{n(n-1)} \sum_{\alpha \neq \beta} s(x_\alpha^{(1)} - x_\beta^{(1)}) s(x_\alpha^{(2)} - x_\beta^{(2)}). \tag{9.9}$$

t will be briefly referred to as the *difference sign covariance* of the sample (9.7).

If all $x^{(1)}$'s and all $x^{(2)}$'s are different, we have

$$\sum_{\alpha \neq \beta} s^2(x_\alpha^{(i)} - x_\beta^{(i)}) = n(n-1), \qquad (i = 1, 2),$$

and then t is the product moment correlation of the difference signs.

It is easily seen that t is a linear function of the number of inversions in the permutation of the ranks of $x^{(1)}$ and $x^{(2)}$.

The statistic t has been considered by Esscher [6], Lindeberg [15], [16], Kendall [12], and others.

t is a U-statistic. As a function of a random sample from a bivariate population, t is an unbiased estimate of the regular functional of degree 2,

$$\tau = \iiiint s(x_1^{(1)} - x_2^{(1)}) s(x_1^{(2)} - x_2^{(2)}) \, dF(x_1) \, dF(x_2). \tag{9.10}$$

τ is the covariance of the signs of differences of the corresponding components of $X_1 = (X_1^{(1)}, X_1^{(2)})$ and $X_2 = (X_2^{(1)}, X_2^{(2)})$ in the population of pairs of independent vectors X_1, X_2 with identical d.f. $F(x) = F(x^{(1)}, x^{(2)})$. If $F(x^{(1)}, x^{(2)})$ is continuous, τ is the product moment correlation of the difference signs.

Two points (or vectors), $(x_1^{(1)}, x_1^{(2)})$ and $(x_2^{(1)}, x_2^{(2)})$ are called concordant or discordant according to whether

$$(x_1^{(1)} - x_2^{(1)})(x_1^{(2)} - x_2^{(2)})$$

is positive or negative. If $\pi^{(c)}$ and $\pi^{(d)}$ are the probabilities that a pair of vectors drawn at random from the population is concordant or discordant, respectively, we have from (9.10)

$$\tau = \pi^{(c)} - \pi^{(d)}.$$

If $F(x^{(1)}, x^{(2)})$ is continuous, we have $\pi^{(c)} + \pi^{(d)} = 1$, and hence

$$\tau = 2\pi^{(c)} - 1 = 1 - 2\pi^{(d)}. \tag{9.11}$$

If we put

$$\bar{F}(x^{(1)}, x^{(2)}) = \tfrac{1}{4}\{F(x^{(1)} - 0, x^{(2)} - 0) + F(x^{(1)} - 0, x^{(2)} + 0) + F(x^{(1)} + 0, x^{(2)} - 0) + F(x^{(1)} + 0, x^{(2)} + 0)\}, \tag{9.12}$$

we have

$$\Phi_1(x \mid \tau) = 1 - 2\bar{F}(x^{(1)}, \infty) - 2\bar{F}(\infty, x^{(2)}) + 4\bar{F}(x^{(1)}, x^{(2)}), \tag{9.13}$$

and we may write

$$\tau = E\{\Phi_1(X_1 \mid \tau)\}. \tag{9.14}$$

The variance of t is, by (5.13),

$$\sigma^2(t) = \frac{2}{n(n-1)}\{2\zeta_1(\tau)(n-2) + \zeta_2(\tau)\}, \tag{9.15}$$

where

$$\zeta_1(\tau) = E\{\Phi_1^2(X_1 \mid \tau)\} - \tau^2, \tag{9.16}$$

$$\zeta_2(\tau) = E\{s^2(X_1^{(1)} - X_2^{(1)})s^2(X_1^{(2)} - X_2^{(2)})\} - \tau^2. \tag{9.17}$$

If $F(x^{(1)}, x^{(2)})$ is continuous, we have $\zeta_2(\tau) = 1 - \tau^2$, and $\bar{F}(x^{(1)}, x^{(2)})$ in (9.13) may be replaced by $F(x^{(1)}, x^{(2)})$.

The variance of a linear function of t has been given for the continuous case by Lindeberg [15], [16].

If $X^{(1)}$ and $X^{(2)}$ are independent and have a continuous d.f., we find $\zeta_1(\tau) = \frac{1}{9}$, $\zeta_2(\tau) = 1$, and hence

$$\sigma^2(t) = \frac{2(2n+5)}{9n(n-1)}. \tag{9.18}$$

In this case the distribution of t is independent of the univariate distributions of $X^{(1)}$ and $X^{(2)}$. This is, however, no longer true if the independent variables are discontinuous. Then it appears that $\sigma^2(t)$ depends on $P\{X_1^{(i)} = X_2^{(i)}\}$ and $P\{X_1^{(i)} = X_2^{(i)} = X_3^{(i)}\}$, $(i = 1, 2)$.

By Theorem 7.1, the d.f. of $\sqrt{n}(t - \tau)$ tends to the normal form. This result has first been obtained for the particular case that all permutations of the ranks of $X^{(1)}$ and $X^{(2)}$ are equally probable, which corresponds to the independence of the continuous random variables $X^{(1)}$, $X^{(2)}$ (Kendall [12]). In this case t can be represented as a sum of independent random variables (cf. Dantzig [5] and Feller [7]). In the general case the asymptotic normality of t has been shown by Daniels and Kendall [4] and the author [10].

The functional $\tau(F)$ is stationary (and hence the normal limiting distribution of $\sqrt{n}(t - \tau)$ singular) if $\zeta_1 = 0$, which, in the case of a continuous F, means that the equation $\Phi_1(X \mid \tau) = \tau$ or

$$4F(X^{(1)}, X^{(2)}) = 2F(X^{(1)}, \infty) + 2F(\infty, X^{(2)}) - 1 + \tau \tag{9.19}$$

is satisfied with probability 1. This is the case if $X^{(2)}$ is an increasing function of $X^{(1)}$. Then $t = \tau = 1$ with probability 1, and $\sigma^2(t) = 0$. A case where (9.19) is fulfilled and $\sigma^2(t) > 0$ is the following: $X^{(1)}$ is uniformly distributed in the interval (0, 1), and

$$X^{(2)} = X^{(1)} + \tfrac{1}{2} \text{ if } 0 \leq X^{(1)} < \tfrac{1}{2},\ X^{(2)} = X^{(1)} - \tfrac{1}{2} \text{ if } \tfrac{1}{2} \leq X^{(1)} \leq 1. \tag{9.20}$$

In this case $\tau = 0, \zeta_2 = 1, \sigma^2(t) = 2/n(n-1)$.

(e) *Rank correlation and grade correlation.* If in the sample $\{(x_\alpha^{(1)}, x_\alpha^{(2)})\}$, $(\alpha = 1, \cdots, n)$, all $x_\alpha^{(1)}$'s and all $x_\alpha^{(2)}$'s are different, the rank correlation coefficient, which we denote by k', is given by

$$k' = \frac{12}{n^3 - n} \sum_{\alpha=1}^{n} \left(R_\alpha^{(1)} - \frac{n+1}{2}\right)\left(R_\alpha^{(2)} - \frac{n+1}{2}\right).$$

Inserting (9.5) we have

$$k' = \frac{3}{n^3 - n} \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \sum_{\gamma=1}^{n} s(x_\alpha^{(1)} - x_\beta^{(1)}) s(x_\alpha^{(2)} - x_\gamma^{(2)})$$

or

$$k' = \frac{(n-2)k + 3t}{n+1} \tag{9. 21}$$

where t is the difference sign covariance (9.9), and

$$k = \frac{3}{n(n-1)(n-2)} \sum{}'' s(x_\alpha^{(1)} - x_\beta^{(1)}) s(x_\alpha^{(2)} - x_\gamma^{(2)}),$$

the summation being over all different subscripts α, β, γ.

k is a U-statistic, and as a function of a random sample from a population with d.f. F, k is an unbiased estimate of the regular functional of degree 3,

$$\begin{aligned} \kappa &= 3 \int \cdots \int s(x_1^{(1)} - x_2^{(1)}) s(x_1^{(2)} - x_3^{(2)})\, dF(x_1)\, dF(x_2)\, dF(x_3) \\ &= 3 \iint \{2\bar{F}^{(1)}(x^{(1)}) - 1\}\{2\bar{F}^{(2)}(x^{(2)}) - 1\}\, dF(x), \end{aligned} \tag{9.22}$$

where $\bar{F}^{(1)}(x^{(1)}) = \bar{F}(x^{(1)}, \infty)$, $\bar{F}^{(2)}(x^{(2)}) = \bar{F}(\infty, x^{(2)})$.

If F is continuous, we have

$$\int \bar{F}^{(i)}(y)\, d\bar{F}^{(i)}(y) = \int_0^1 u\, du = \tfrac{1}{2},$$

$$\int \{\bar{F}^{(i)}(y) - \tfrac{1}{2}\}^2\, d\bar{F}^{(i)}(u) = \int_0^1 (u - \tfrac{1}{2})^2\, du = \tfrac{1}{12}, \qquad (i = 1, 2),$$

and in this case κ is the coefficient of correlation between the random variables

$$U^{(1)} = F^{(1)}(X^{(1)}), \qquad U^{(2)} = F^{(2)}(X^{(2)}).$$

$U^{(i)}$ has been termed the grade of the continuous variable $X^{(i)}$, and in the general case $\bar{F}^{(i)}(X^{(i)})$ may be called the grade of $X^{(i)}$ (cf., for instance, G. U. Yule and M. G. Kendall [22, p. 150]). In general, κ is 12 times the covariance of the grades.

From (9.21) we have for the expected value of k',

$$E\{k'\} = \frac{(n-2)\kappa + 3\tau}{n+1}.$$

In the continuous case the rank correlation coefficient k' is an estimate of the grade correlation κ, which is biased for finite n but unbiased in the limit.

The kernel $3s(x_1^{(1)} - x_2^{(1)})s(x_1^{(2)} - x_3^{(2)})$ of κ is not symmetric. Denoting by $\Phi(x_1, x_2, x_3 \mid \kappa)$ the symmetric kernel of κ, we have

$$\Phi(x_1, x_2, x_3 \mid \kappa) = \frac{1}{2} \sum_{\substack{\alpha \neq \beta \neq \gamma \\ \alpha \neq \gamma}}^{1,2,3} s(x_\alpha^{(1)} - x_\beta^{(1)})s(x_\alpha^{(2)} - x_\gamma^{(2)}) \tag{9.23}$$

For computing κ and the constants ζ_i an alternative expression for κ and Φ is sometimes more convenient. From three two-dimensional vectors x_1, x_2, x_3 we can form three pairs (x_1, x_2), (x_1, x_3), and (x_2, x_3). The number of concordant pairs among them can be 3, 2, 1, or 0. If γ is the probability that among the three pairs formed from three random elements of the population at least 2 are concordant, we have, if the d.f. F is continuous,

$$\kappa = 2\gamma - 1. \tag{9.24}$$

This is analogous to the expression (9.11) for τ.

The truth of (9.24) can be seen as follows: From the definition of γ we have

$$\gamma = E\{\Phi(x_1, x_2, x_3 \mid \gamma)\},$$

where $\Phi(x_1, x_2, x_3 \mid \gamma)$ is $= 1$ if at least two of the three expressions

$$(x_\alpha^{(1)} - x_\beta^{(1)})(x_\alpha^{(2)} - x_\beta^{(2)}), \qquad (\alpha < \beta;\ \alpha, \beta = 1, 2, 3) \tag{9.25}$$

are positive, and equal to zero, if no more than one of them is positive. Since, by the continuity of F, we may neglect the case of (9.25) being zero, we may write

$$\Phi(x_1, x_2, x_3 \mid \gamma) = c_{12,12}c_{23,23}c_{31,31} + c_{12,12}c_{23,23}c_{31,13} + c_{12,12}c_{23,32}c_{31,31} + c_{12,21}c_{23,23}c_{31,31},$$

where

$$c_{\alpha,\beta,\gamma,\delta} = c[(x_\alpha^{(1)} - x_\beta^{(1)})(x_\gamma^{(2)} - x_\delta^{(2)})]$$

and $c(u)$ is defined by (9.4).

$\Phi(x_1, x_2, x_3 \mid \gamma)$ is symmetric in x_1, x_2, x_3.

The identity

$$\Phi(x_1, x_2, x_3 \mid \kappa) = 2\Phi(x_1, x_2, x_3 \mid \gamma) - 1 \tag{9.26}$$

can be shown to hold either by algebraical calculation using (9.4) or by direct computation of each side for the different positions of the three points x_1, x_2, x_3.

From (9.26) it appears that in the continuous case the symmetric kernel $\Phi(x_1, x_2, x_3 \mid \kappa)$ can assume only two values, -1 and $+1$.

The variance of k is, according to (5.13),

$$\sigma^2(k) = \frac{6}{n(n-1)(n-2)} \left\{ 3\binom{n-3}{2} \zeta_1(\kappa) + 3(n-3)\zeta_2(\kappa) + \zeta_3(\kappa) \right\},$$

where

$$\zeta_1(\kappa) = E\{\Phi_1^2(X_1 \mid \kappa)\} - \kappa^2,$$

$$\zeta_2(\kappa) = E\{\Phi_2^2(X_1, X_2 \mid \kappa)\} - \kappa^2,$$

$$\zeta_3(\kappa) = E\{\Phi^2(X_1, X_2, X_3 \mid \kappa)\} - \kappa^2,$$

$$\Phi_1(x_1 \mid \kappa) = E\{\Phi(x_1, X_2, X_3 \mid \kappa)\},$$

$$\Phi_2(x_1, x_2 \mid \kappa) = E\{\Phi(x_1, x_2, X_3 \mid \kappa)\}.$$

We find for the continuous case

$$\zeta_3(\kappa) = 1 - \kappa^2,$$

$$(9.27) \quad \begin{aligned} \Phi_1(x_1 \mid \kappa) = {} & [1 - 2F(x_1^{(1)}, \infty)][1 - 2F(\infty, x_1^{(2)})] - 2F(x_1^{(1)}, \infty) \\ & - 2F(\infty, x_1^{(2)}) + 4\int F(x_1^{(1)}, y^{(2)})dF(\infty, y^{(2)}) \\ & + 4\int F(y^{(1)}, x_1^{(2)})dF(y^{(1)}, \infty), \end{aligned}$$

$$\begin{aligned} \Phi_2(x_1, x_2 \mid \kappa) = {} & 1 + 2F(x_1^{(1)}, x_2^{(2)}) + 2F(x_2^{(1)}, x_1^{(2)}) - 2c(x_2^{(2)} - x_1^{(2)})F(x_1^{(1)}, \infty) \\ & - 2c(x_1^{(2)} - x_2^{(2)})F(x_2^{(1)}, \infty) - 2c(x_2^{(1)} - x_1^{(1)})F(\infty, x_1^{(2)}) \\ & - 2c(x_1^{(1)} - x_2^{(1)})F(\infty, x_2^{(2)}). \end{aligned}$$

If $X^{(1)}$, $X^{(2)}$ are continuous and independent, we obtain $\kappa = 0$, $\zeta_1 = \frac{1}{9}$, $\zeta_2 = \frac{7}{18}$, $\zeta_3 = 1$, and hence

$$(9.28) \quad \sigma^2(k) = \frac{n^2 - 3}{n(n-1)(n-2)}.$$

In the discontinuous case of independence the distribution of k, as that of t, depends on the distributions of $X^{(1)}$ and $X^{(2)}$, and $\sigma^2(k)$ can again be expressed in terms of $P\{X_1^{(i)} = X_2^{(i)}\}$ and $P\{X_1^{(i)} = X_2^{(i)} = X_3^{(i)}\}$, $(i = 1, 2)$.

The variance of the rank correlation coefficient k' is, by (9.21),

$$(9.29) \quad \sigma^2(k') = \frac{(n-2)^2\sigma^2(k) + 6(n-2)\sigma(t, k) + 9\sigma^2(t)}{(n+1)^2}.$$

For $\sigma(t, k)$ we have, according to (6.5),

$$\sigma(t, k) = \frac{6}{n(n-1)} \{(n-3)\zeta_1(\tau, \kappa) + \zeta_2(\tau, \kappa)\},$$

where

$$\zeta_1(\tau, \kappa) = E\{\Phi_1(X_1 \mid \tau)\Phi_1(X_1 \mid \kappa)\} - \tau\kappa,$$

$$\zeta_2(\tau, \kappa) = E\{\Phi(X_1, X_2 \mid \tau)\Phi_2(X_1, X_2 \mid \kappa)\} - \tau\kappa.$$

In the case of independence we see from (9.13) and (9.27) that

$$\Phi_1(x \mid \tau) = \Phi_1(x \mid \kappa) = [1 - 2F(x^{(1)}, \infty)][1 - 2F(\infty, x^{(2)})],$$

and we obtain

$$\zeta_1(\tau, \kappa) = \zeta_1(\kappa) = \zeta_1(\tau) = \tfrac{1}{9}, \tag{9.30}$$

$$\zeta_2(\tau, \kappa) = \tfrac{5}{9},$$

$$\sigma(t, k) = \frac{2(n+2)}{3n(n-1)}. \tag{9.31}$$

On inserting (9.28), (9.31) and (9.18) in (9.29), we find

$$\sigma^2(k') = \frac{1}{n-1},$$

in accordance with the result obtained for this case by Student and published by K. Pearson [20].

According to Theorem 7.1, $\sqrt{n}(k - \kappa)$ tends to be normally distributed with mean 0 and variance $9\zeta_1(\kappa)$. The same is true for the distribution of the rank correlation coefficient, k', as follows from Theorem 7.3 in conjunction with (9.21). For the special case of independence the asymptotic normality of k' has been proved by Hotelling and Pabst [11].

From Theorem 7.3 it also follows that the joint distribution of $\sqrt{n}(t - \tau)$ and $\sqrt{n}(k - \kappa)$ (or $\sqrt{n}(k' - \kappa)$) tends to the normal form with the variances $4\zeta_1(\tau)$ and $9\zeta_1(\kappa)$ and the covariance $6\zeta_1(\kappa, \tau)$. In the case of independence we see from (9.30) that the correlation $\rho(t, k)$ between t and k tends to 1, and we have the asymptotic functional relation $3t = 2k$. This result has been conjectured by Kendall and others [14], and proved by Daniels [3]. In general, however, $\rho(t, k)$ does not approach unity. Thus, if $X^{(1)}$ is uniformly distributed in (0, 1), and

$$\begin{aligned}
X^{(2)} &= \tfrac{1}{2} - X^{(1)} && \text{if } 0 \le X^{(1)} < \tfrac{1}{4}, \\
X^{(2)} &= \tfrac{1}{2} + X^{(1)} && \text{if } \tfrac{1}{4} \le X^{(1)} < \tfrac{1}{2}, \\
X^{(2)} &= X^{(1)} - \tfrac{1}{2} && \text{if } \tfrac{1}{2} \le X^{(1)} < \tfrac{3}{4}, \\
X^{(2)} &= \tfrac{3}{2} - X^{(1)} && \text{if } \tfrac{3}{4} \le X^{(1)} \le 1,
\end{aligned} \tag{9.32}$$

we have $\tau = \kappa = 0$, $\zeta_1(\tau) = 0$, $\zeta_2(\tau) = 1$, $\zeta_1(\kappa) = \frac{1}{16}$, $\zeta_1(\kappa, \tau) = 0$, and hence $\rho(t, k) \to 0$.

(f) *Non-parametric tests of independence.* Suppose that the random variables $X^{(1)}$, $X^{(2)}$ have a continuous joint d.f. $F(x^{(1)}, x^{(2)})$, and we want to test the hypothesis H_0 that $X^{(1)}$ and $X^{(2)}$ are independent, that is, that

$$F(x^{(1)}, x^{(2)}) = F(x^{(1)}, \infty)\, F(\infty, x^{(2)}).$$

The distribution of any statistic involving only the ranks of the variables does not depend on the d.f. of the population when H_0 is true. For this reason several rank order statistics, among them the difference sign correlation t and the rank correlation k', have been suggested for testing independence.

From the preceding results we can obtain the asymptotic power functions of the tests of independence based on t and k'. If H_0 is true, we have $E\{t\} = \tau = 0$, and the critical region of size ϵ of the t-test may be defined by $|t| > c_n$, where c_n is the smallest number satisfying the inequality

$$P\{\, |t| > c_n \mid H_0\} \leq \epsilon. \tag{9.33}$$

By Theorem 7.2 and (9.18) we may write $c_n = 2\lambda_n/3\sqrt{n}$, where λ_n tends to a positive constant λ depending on ϵ.

Since $\sigma^2(t) = O(n^{-1})$, the power function

$$P_n(H) = P\{\, |t| \geq 2\lambda_n/3\sqrt{n} \mid H\}$$

tends to one as $n \to \infty$ for any alternative hypothesis H with $\tau(F) \neq 0$. If, however, $\tau = 0$, we have $\lim P_n(H) < 1$. If $\tau = 0$ and $\zeta_1(\tau) < \frac{1}{9}$, we have even $\lim P_n(H) < \epsilon$, and with respect to these alternatives the test is biased in the limit. Thus, in the case of the distribution (9.20) we have even $P_n(H) \to 0$. In this case there is a functional relationship between the variables, and the distribution must be considered as considerably different from the case of independence.

For the rank correlation test we have a similar result. If c_n' is the smallest number satisfying $P\{\, |k'| > c_n' \mid \mathrm{H}_0\} \leq \epsilon$, we have $c_n' = \lambda_n'/\sqrt{n}$, where $\lim \lambda_n' = \lambda$, and the test is biased in the limit if $\kappa = 0$ and $\zeta_1(\kappa) < \frac{1}{9}$. This is fulfilled in the case of the distribution (9.32), where $\zeta_1(\kappa) = \frac{1}{16}$.

The question arises whether there exist non-parametric tests of independence which are unbiased or unbiased in the limit. This point will be discussed in a separate paper on tests of independence.

(g) *Mann's test against trend.* Let $Y_1, \cdots, Y_n$ be n independent real-valued random variables, Y_α having the continuous d.f. $F_\alpha(y)$, $(\alpha = 1, \cdots, n)$. The hypothesis of randomness,

$$H_1: \quad F_1(y) = \cdots = F_n(y)$$

is to be tested against the alternative hypothesis of a "downward trend,"

$$H_2: \quad F_1(y) < F_2(y) < \cdots < F_n(y).$$

H. B. Mann [17] has suggested a test of H_1 against H_2 based on the number T of inequalities $Y_\alpha < Y_\beta$, where $\alpha < \beta$. We may write

$$2T - \frac{n(n-1)}{2} = \sum_{\alpha<\beta} s(Y_\beta - Y_\alpha) = \sum_{\alpha<\beta} s(\alpha - \beta)s(Y_\alpha - Y_\beta).$$

The U-statistic

$$t = \{4T/n(n-1)\} - 1$$

is the same as (9.9) for the special case when one component is not a random variable.

Let

$$\tau_{\alpha\beta} = s(\alpha - \beta) \int\int s(y_1 - y_2)\, dF_\alpha(y_1)\, dF_\beta(y_2)$$
$$= s(\alpha - \beta)\left\{2 \int F_\beta(y)\, dF_\alpha(y) - 1\right\}.$$

We have $\tau_{\alpha\beta} = 0$ if H_1 is true and $\tau_{\alpha\beta} < 0$ if H_2 is true.

Since

$$E\{t\} = \tau_n = \frac{2}{n(n-1)} \sum_{\alpha<\beta} \tau_{\alpha\beta},$$

it follows that $E\{t\} = 0$ under H_1 and $E\{t\} < 0$ under H_2.

Mann's test against trend has the power function $P_n(H) = P\{t < a_n \mid H\}$, where a_n is the largest number satisfying $P\{t < a_n \mid H_1\} \leq \epsilon$.

Since $a_n \to 0$ and, by (5.18), $\sigma^2(t) = O(n^{-1})$, it follows from Tchebycheff's inequality that the test is consistent (that is, $P_n(H_2) \to 1$) and hence unbiased in the limit. This has been shown by Mann who also gave sufficient conditions under which the test is unbiased for finite n.

By Theorems 8.1 and 8.2 the distribution of $(t - \tau_n)/\sigma(t)$ is asymptotically normal if certain conditions are satisfied. Since (8.2), (8.3) and (8.13) are fulfilled, either of the conditions (8.4) and (8.14) is sufficient.

(h) *The coefficient of partial difference sign correlation.* Consider a three-variate sample $x_1, \cdots, x_n$; $x_\alpha = (x_\alpha^{(1)}, x_\alpha^{(2)}, x_\alpha^{(3)})$, $(\alpha = 1, \cdots, n)$. In a similar way as in section 9d we may form the set of the $n(n-1)$ triplets of difference signs,

$$s(x_\alpha^{(1)} - x_\beta^{(1)}), \quad s(x_\alpha^{(2)} - x_\beta^{(2)}), \quad s(x_\alpha^{(3)} - x_\beta^{(3)}), \qquad (\alpha \neq \beta;\ \alpha, \beta = 1, \cdots, n). \tag{9.34}$$

We shall assume that all $x^{(1)}$'s, $x^{(2)}$'s, and $x^{(3)}$'s are different. Then the triplets (9.34) contain only two different numbers, $+1$ and -1. Hence the regression functions of the three-variate population (9.34) are linear.

If t_{12}, t_{13}, and t_{23} are the difference sign correlations of $\{s(x_\alpha^{(1)} - x_\beta^{(1)}), s(x_\alpha^{(2)} - x_\beta^{(2)})\}$, $\{s(x_\alpha^{(1)} - x_\beta^{(1)}), s(x_\alpha^{(3)} - x_\beta^{(3)})\}$ and $\{s(x_\alpha^{(2)} - x_\beta^{(2)}), s(x_\alpha^{(3)} - x_\beta^{(3)})\}$ respectively, we have for the coefficient $t_{12.3}$ of partial correlation between $s(x_\alpha^{(1)} - x_\beta^{(1)})$ and $s(x_\alpha^{(2)} - x_\beta^{(2)})$ with respect to $s(x_\alpha^{(3)} - x_\beta^{(3)})$,

$$t_{12.3} = \frac{t_{12} - t_{13}t_{23}}{\sqrt{(1 - t_{13}^2)(1 - t_{23}^2)}}. \tag{9.35}$$

This measure of partial correlation has been suggested by Kendall [13] who gave an alternative definition of $t_{12.3}$.

If we have two independent three-dimensional random vectors $X_1 = (X_1^{(1)}, X_1^{(2)}, X_1^{(3)})$ and $X_2 = (X_2^{(1)}, X_2^{(2)}, X_2^{(3)})$ with the same continuous d.f. $F(x^{(1)}, x^{(2)}, x^{(3)})$, the distribution of the difference signs $s(X_1^{(i)} - X_2^{(i)})$, $(i = 1, 2, 3)$, has again linear regression functions, and we may define the partial difference sign correlation

$$\tau_{12.3} = \frac{\tau_{12} - \tau_{13}\tau_{23}}{\sqrt{(1 - \tau_{13}^2)(1 - \tau_{23}^2)}},$$

where τ_{ij} is the difference sign correlation of $X^{(i)}, X^{(j)}$.

If $t_{12.3}$ is a function of a random sample, and if $\tau_{13}^2 \neq 1$, $\tau_{23}^2 \neq 1$, the d.f. of $\sqrt{n}(t_{12.3} - \tau_{12.3})$ tends, by Theorem 7.5, so the normal d.f. with mean zero and variance

$$\sigma_{12.3}^2 = \frac{4}{(1 - \tau_{13}^2)(1 - \tau_{23}^2)} \left\{ \zeta_1(\tau_{12}) + \frac{(\tau_{23} - \tau_{12}\tau_{13})^2}{(1 - \tau_{13}^2)^2} \zeta_1(\tau_{13}) \right.$$
$$+ \frac{(\tau_{13} - \tau_{12}\tau_{23})^2}{(1 - \tau_{23}^2)^2} \zeta_1(\tau_{23}) - 2\,\frac{\tau_{23} - \tau_{12}\tau_{13}}{1 - \tau_{13}^2} \zeta_1(\tau_{12}, \tau_{13}) - 2\,\frac{\tau_{13} - \tau_{12}\tau_{23}}{1 - \tau_{23}^2} \zeta_1(\tau_{12}, \tau_{23})$$
$$\left. + 2\,\frac{(\tau_{23} - \tau_{12}\tau_{13})(\tau_{13} - \tau_{12}\tau_{23})}{(1 - \tau_{13}^2)(1 - \tau_{23}^2)} \zeta_1(\tau_{13}, \tau_{23}) \right\},$$

where

$$\zeta(\tau_{ij}) = E\,\{\Phi_1^2(X \mid \tau_{ij})\} - \tau_{ij}^2,$$
$$\zeta_1(\tau_{ij}, \tau_{gh}) = E\{\Phi_1(X \mid \tau_{ij})\Phi_1(X \mid \tau_{gh})\} - \tau_{ij}\tau_{gh},$$

and, for instance (cf. (9.13)),

$$\Phi_1(x \mid \tau_{12}) = 1 - 2F(x^{(1)}, \infty, \infty) - 2F(\infty, x^{(2)}, \infty) + 4F(x^{(1)}, x^{(2)}, \infty).$$

If $\tau_{13} = \tau_{23} = 0$, we have

$$\sigma_{12.3}^2 = 4\zeta_1(\tau_{12}),$$

and $\sqrt{n}(t_{12.3} - \tau_{12.3})$ has the same limiting distribution as $\sqrt{n}(t_{12} - \tau_{12})$. This is in particular the case when $X^{(1)}, X^{(2)}, X^{(3)}$ are independent.

REFERENCES

[1] H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tracts in Math., Cambridge, 1937.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[3] H. E. Daniels, "The relation between measures of correlation in the universe of sample permutations," *Biometrika*, Vol. 33 (1944), pp. 129–135.

[4] H. E. Daniels and M. G. Kendall, "The significance of rank correlations where parental correlation exists," *Biometrika*, Vol. 34 (1947), pp. 197–208.

[5] G. B. Dantzig, "On a class of distributions that approach the normal distribution function," *Annals of Math. Stat.*, Vol. 10 (1939) pp. 247–253.

[6] F. Esscher, "On a method of determining correlation from the ranks of the variates," *Skandinavisk Aktuar. tids.*, Vol. 7 (1924), pp. 201–219.

[7] W. Feller, "The fundamental limit theorems in probability," *Am. Math. Soc. Bull.*, Vol. 51 (1945), pp. 800–832.

[8] C. Gini, "Sulla misura della concentrazione e della variabilità dei caratteri," *Atti del R. Istituto Veneto di S.L.A.*, Vol. 73 (1913–14), Part 2.

[9] P. R. Halmos, "The theory of unbiased estimation," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 34–43.

[10] W. Höffding, "On the distribution of the rank correlation coefficient τ, when the variates are not independent," *Biometrika*, Vol. 34 (1947), pp. 183–196.

[11] H. Hotelling and M. R. Pabst, "Rank correlation and tests of significance involving no assumptions of normality," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 29–43.

[12] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, Vol. 30 (1938), pp. 81–93.

[13] M. G. Kendall, "Partial rank correlation," *Biometrika*, Vol. 32 (1942), pp. 277–283.

[14] M. G. Kendall, S. F. H. Kendall, and B. Babington Smith, "The distribution of Spearman's coefficient of rank correlation in a universe in which all rankings occur an equal number of times," *Biometrika*, Vol. 30 (1939), pp. 251–273.

[15] J. W. Lindeberg, "Über die Korrelation," VI *Skand. Matematikerkongres i København*, 1925, pp. 437–446.

[16] J. W. Lindeberg, "Some remarks on the mean error of the percentage of correlation," *Nordic Statistical Journal*, Vol. 1 (1929), pp. 137–141.

[17] H. B. Mann, "Nonparametric tests against trend," *Econometrica*, Vol. 13 (1945), pp. 245–259.

[18] R. v. Mises, "On the asymptotic distribution of differentiable statistical functions," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 309–348.

[19] U. S. Nair, "The standard error of Gini's mean difference," *Biometrika*, Vol. 28 (1936), 428–436.

[20] K. Pearson, "On further methods of determining correlation," *Drapers' Company Research Memoirs*, Biometric Series, IV, London, 1907.

[21] V. Volterra, *Theory of Functionals*, Blackie, (authorized translation by Miss M. Long), London and Glasgow, 1931.

[22] G. U. Yule and M. G. Kendall, *An Introduction to the Theory of Statistics*, Griffin, 11th Edition, London, 1937.

OPTIMUM CHARACTER OF THE SEQUENTIAL PROBABILITY RATIO TEST

A. Wald and J. Wolfowitz

Columbia University

1. Summary. Let S_0 be any sequential probability ratio test for deciding between two simple alternatives H_0 and H_1, and S_1 another test for the same purpose. We define $(i, j = 0, 1)$:

$\alpha_i(S_j)$ = probability, under S_j, of rejecting H_i when it is true;

$E_i^j(n)$ = expected number of observations to reach a decision under test S_j when the hypothesis H_i is true. (It is assumed that $E_i^1(n)$ exists.)

In this paper it is proved that, if

$$\alpha_i(S_1) \leq \alpha_i(S_0) \qquad (i = 0,1),$$

it follows that

$$E_i^0(n) \leq E_i^1(n) \qquad (i = 0, 1).$$

This means that of all tests with the same power the sequential probability ratio test requires on the average fewest observations. This result had been conjectured earlier ([1], [2]).

2. Introduction. Let $p_i(x)$, $i = 0, 1$, denote two different probability density functions or (discrete) probability functions. (Throughout this paper the index i will always take the values 0, 1). Let X be a chance variable whose distribution can only be either $p_0(x)$ or $p_1(x)$, but is otherwise unknown. It is required to decide between the hypotheses H_0, H_1, where H_i states that $p_i(x)$ is the distribution of X, on the basis of n independent observations $x_1, \cdots, x_n$ on X, where n is a chance variable defined (finite) on almost every infinite sequence

$$\omega = x_1, x_2, \cdots$$

i.e., n is finite with probability one according to both $p_0(x)$ and $p_1(x)$. The definition of $n(\omega)$ together with the rule for deciding on H_0 or H_1 constitute a sequential test.

A sequential probability ratio test is defined with the aid of two positive numbers, $A^* > 1$, $B^* < 1$, as follows: Write for brevity

$$p_{ij} = \prod_{k=1}^{j} p_i(x_k).$$

Then $n = j$ if

$$\frac{p_{1j}}{p_{0j}} \geq A^* \qquad \text{or} \qquad \leq B^*$$

and

$$B^* < \frac{p_{1k}}{p_{0k}} < A^*, \qquad k < j.$$

If

$$\frac{p_{1n}}{p_{0n}} \geq A^*, \qquad \text{the hypothesis } H_1 \text{ is accepted,}$$

if

$$\frac{p_{1n}}{p_{0n}} \leq B^* \text{ the hypothesis } H_0 \text{ is accepted.}$$

In this paper we limit consideration to sequential tests for which $E_i(n)$ exists, where $E_i(n)$ is the expected value of n when H_i is true (i.e., when $p_i(x)$ is the distribution of X). It has been proved in [3] that all sequential probability ratio tests belong to this class. The purpose of the paper is to prove the result stated in the first section. Throughout the proof we shall find it convenient to assume that there is an a priori probability g_i that H_i is true ($g_0 + g_1 = 1$; we shall write $g = (g_0, g_1)$). We are aware of the fact that many statisticians believe that in most problems of practical importance either no a priori probability distribution exists, or that even where it exists the statistical decision must be made in ignorance of it; in fact we share this view. Our introduction of the a priori probability distribution is a purely technical device for achieving the proof which has no bearing on statistical methodology, and the reader will verify that this is so. We shall always assume below that $g_0 \neq 0, 1$.

Let W_0, W_1, c be given positive numbers. We define

$$R = g_0(W_0\alpha_0 + cE_0(n)) + g_1(W_1\alpha_1 + cE_1(n)),$$

and call R the average risk associated with a test S and a given g (obviously R is a function of both). We shall say that H_i is accepted when the decision is made that $p_i(x)$ is the distribution of X. We shall say that H_0 is rejected when H_1 is accepted, and vice versa. The reader may find it helpful to regard W_i as a weight which measures the loss caused by rejecting H_i when it is true, c as the cost of a single observation, and R as the average loss associated with a given g and a test S. For mathematical purposes these are simply quantities which we manipulate in the course of the proof.

3. Role of the probability ratio. Let g, $W = (W_0, W_1)$, and c be fixed. Let S be a given sequential test, with $R(S)$ the associated risk and $n(\omega, S)$ the associated "sample size" function. Let $\psi(x_1, \cdots, x_n)$ be the "decision" function; this is a function which takes only the values 0 and 1, and such that, when $x_1, \cdots, x_n$ is the sample point, the hypothesis with index $\psi(x_1, \cdots, x_n)$ is *rejected*. Define the following decision function $\varphi(x_1, \cdots, x_n)$: $\varphi = 0$ when

$$\lambda = \frac{W_1 g_1 p_{1n}}{W_0 g_0 p_{0n}}$$

is greater than 1, and $\varphi = 1$ when $\lambda < 1$. When $\lambda = 1$, φ may be 0 or 1 at pleasure.

It must be remembered that an actual decision function is a single-valued function of $(x_1, \cdots, x_n)$. We note, however, that

a) the relevant properties of a test are not affected by changing the test on a set T of points ω whose probability is zero according to both H_0 and H_1, i.e., changing the definition on T of n and/or of the decision function, leaves α_0, α_1, $E_0(n)$ and $E_1(n)$ unaltered. In particular, the average risk R remains unchanged.

b) the set of points for which $p_{0n} = p_{1n} = 0$ and λ is indeterminate, has probability zero according to both H_0 and H_1.

In view of the above we decide arbitrarily, in all sequential tests which we shall henceforth consider, to define $n = j$, and $\psi = 0$, whenever $p_{0j} = p_{1j} = 0$, and $n \neq 1, \cdots, (j-1)$. By this arbitrary action $R(S)$ will not be changed.

Let now

$$L_{in} = \frac{W_i g_i p_{in}}{g_0 p_{0n} + g_1 p_{1n}};$$

$$L_n = cn + \min (L_{0n}, L_{1n}).$$

We have

$$EL_{\psi n} = \Sigma g_i W_i \alpha_i$$

where the operator E denotes the expected value with respect to the joint distribution of H_i and $(x_1, \cdots, x_n)$, i.e., E is the operator $g_0 E_0 + g_1 E_1$. If now the event $\{\psi(S) \neq \varphi \text{ and } \lambda \neq 1\}$ has positive probability according to either H_0 or H_1, we would have, for $n = n(\omega, S)$,

$$EL_{\varphi n} < EL_{\psi n}.$$

Hence, if the decision function ψ connected with the test S were replaced by the decision function φ, R would be decreased. Since our object throughout this proof will be to make R as small as possible, we shall confine ourselves henceforth, except when the contrary is explicitly stated, to tests for which φ is the decision function. This will be assumed even if not explicitly stated.

The function φ has not yet been uniquely defined when $\lambda = 1$. A definition convenient for later purposes will be given in the next section. R is the same for all definitions.

We thus have that φ is a function only of λ, or, what comes to the same thing when W is fixed, of $r_n = \frac{p_{1n}}{p_{0n}}$. Define

$$r_j = \frac{p_{1j}}{p_{0j}}, \qquad j = 1, 2, \cdots.$$

We shall now prove

LEMMA 1. *Let g, W, and c be fixed. There exists a sequential test S^* for which the average risk is a minimum. Its sample size function $n(\omega, S^*)$ can be defined by means of a properly chosen subset K of the non-negative half-line as follows: For any ω consider the associated sequence*

$$r_1, r_2, \cdots$$

and let j be the smallest integer for which $r_j \epsilon K$. Then $n = j$. The function n may be undefined on a set of points ω whose probability according to H_0 and H_1 is zero.

Let $a = (a_1, \cdots, a_d)$ be any point in some finite d-dimensional Euclidean space, provided only that $p_{0d}(a)$ and $p_{1d}(a)$ are not both zero. Let $b = \frac{p_{1d}(a)}{p_{0d}(a)}$ and let $l(a) = cd + \min(L_{0d}, L_{1d})$. Let D be any sequential test whatever for which $n(\omega, D) > d$ for any ω whose first d coördinates are the same as those of a, and for which $E(n \mid a, D) < \infty$, where $E(n \mid a, D)$ is the conditional expected value of n according to the test D under the condition that the first d coördinates of ω are the same as those of a. For brevity let G represent the set of points ω which fulfill this last condition, i.e., that the first d coördinates of ω are the same as those of a. Finally, let $E(L_n \mid a, D)$ be the conditional expected value of L_n according to D under the condition that ω is in the set G. We know that $\min(L_{0d}, L_{1d})$ depends only on $r_d(a) = b$.

Write

$$\nu(a) = \sup_D [l(a) - E(L_n \mid a, D)].$$

Let $a_0 = (a_{01}, \cdots, a_{0k})$ be any point such that

$$\frac{p_{1d}(a)}{p_{0d}(a)} = \frac{p_{1k}(a_0)}{p_{0k}(a_0)}.$$

Let D_0 be any sequential test whatever for which $n(\omega, D_0) > k$ for any ω whose first k coordinates are the same as those of a_0, and for which $E(n \mid a_0, D_0) < \infty$ Let

$$\nu(a_0) = \sup_{D_0} [l(a_0) - E(L_n \mid a_0, D_0)].$$

We shall prove that $\nu(a) = \nu(a_0)$. Thus we shall be justified in writing

$$\gamma(b) = \nu(a) = \nu(a_0).$$

Suppose, therefore that $\nu(a) > \nu(a_0)$. Let D_1 be a test of the type D such that

$$l(a) - E(L_n \mid a, D_1) > \frac{\nu(a) + \nu(a_0)}{2}.$$

We now partially define another sequential test D_{10} of the type D_0 as follows: Let

$$\bar{a} = a_1, \cdots, a_d, y_1, \cdots, y_t,$$

be any sequence such that $n(\bar{a}, D_1) = d + t$. Then for the sequence

$$\bar{a}_0 = a_{01}, \cdots, a_{0k}, y_1, \cdots, y_t,$$

let $n(\bar{a}_0, D_{10}) = k + t$. The decision function ψ_0 associated with D_{10} will be partially defined as follows:

$$\psi_0(\bar{a}_0) = \varphi(\bar{a}).$$

(The reader will observe that it may happen that $\psi_0(\bar{a}_0) \neq \varphi(\bar{a}_0)$). Since $r_d(a) = r_k(a_0)$ it follows that

$$l(a) - E(L_n \mid a, D_1) = l(a_0) - E(L_n \mid a_0, D_{10}) > \frac{\nu(a) + \nu(a_0)}{2} > \nu(a_0),$$

in violation of the definition of $\nu(a_0)$. A similar contradiction is obtained if $\nu(a) < \nu(a_0)$. Hence $\nu(a) = \nu(a_0)$ as was stated above.

We define K to consist of all numbers b which are such that there exist points a with $r_d(a) = b$, and for which $\gamma(b) \leq 0$. We shall now prove that the test S^* defined in the statement of the lemma is such that $R(S^*)$ is a minimum. Recall that the average risk is the expected value of L_n. Let S be any other test. Let $a^* = (a_1^*, \cdots, a_{d^*}^*)$ be any sequence such that either $n(a^*, S^*) = d^*$, or $n(a^*, S) = d^*$, but $n(a^*, S^*) \neq n(a^*, S)$. We exclude the trivial case that the probability of the occurrence of such a sequence, under both H_0 and H_1, is zero. Let $r_{d^*}(a^*) = b^*$. The sequence a^* may be one of three types:

1) $\gamma(b^*) < 0$. Hence $b^* \in K$, $n(a^*, S) > d^*$. It is more advantageous, from the point of view of diminishing the average risk, to terminate the sequential process at once, since $E(L_n \mid a^*, S) > l(a^*)$.

2) $\gamma(b^*) = 0$. Hence $b^* \in K$, $n(a^*, S) > d^*$. If $l(a^*) - E(L_n \mid a^*, S) = 0$, i.e., the supremum is actually attained by S, then, as far as the average risk is concerned, it makes no difference whether the sequential process is terminated with a^* or continued according to S. If, however, $l(a^*) - E(L_n \mid a^*, S) < 0$, it is clearly disadvantageous to proceed according to S. It is impossible that $l(a^*) - E(L_n \mid a^*, S) > 0$, since $\gamma(b^*) = 0$.

3) $\gamma(b^*) > 0$. Hence $b^* \notin K$, $n(a^*, S) = d^*$. Clearly it is more advantageous from the point of view of diminishing the average risk not to terminate the sequential process, but to continue with at least one more observation. After one more observation we are either in case 1 or 2, where it is advantageous to terminate the sequential process, or again in case 3, where it is advantageous to take yet another observation.

We conclude that $R(S^*)$ is a minimum, as was to be proved.

4. A fundamental lemma. Consider the complement of K with respect to the non-negative half-line, and from it delete all points b' for which there exists no point a in some d-dimensional Euclidean space such that $r_d(a) = b'$. The point 1 is never to be considered as of the type of b', i.e., 1 is never to be deleted. Designate the resulting set by $\bar{K}$.

Our proof of the theorem to which this paper is devoted hinges on the following lemma:

LEMMA 2. *Let W, g, c be fixed, and $\bar{K}$ be as defined above. There exist two positive numbers A and B, with $B \leq \frac{W_0 g_0}{W_1 g_1} \leq A$, such that*

a) if $b \in K$, then either $b \geq A$ or $b \leq B$

b) if $b \in \bar{K}$, $B \leq b \leq A$.

Two remarks may be made before proceeding with the proof:

1) We may now complete the definition of φ for tests of the type of S^*. The reader will recall that φ was not uniquely defined when $\lambda = 1$, i.e., when $r_n = \frac{W_0 g_0}{W_1 g_1}$. Lemma 2 shows that it is necessary to define $\varphi(\lambda)$ only when $\lambda = \frac{W_0 g_0}{W_1 g_1} \in K$ and λ is therefore either A or B. We will define $\varphi\left(\frac{W_0 g_0}{W_1 g_1}\right)$ as 0 or 1, according as $\frac{W_0 g_0}{W_1 g_1}$ is A or B, and $A \neq B$. This is simply a convenient definition which will give uniqueness. When $A = B = \frac{W_0 g_0}{W_1 g_1} \in K$, the situation is completely trivial, and we may take $\varphi = 0$ arbitrarily.

2) If $1 \in K$ the above lemma shows that the average risk is minimized (for fixed W, g, c, of course) by taking no observations at all. We have $\varphi = 0$ or 1 according as $1 \geq A$ or $1 \leq B$.

PROOF OF THE LEMMA: Let $h > \frac{W_0 g_0}{W_1 g_1}$ be a point in $\bar{K}$. We will prove that any point h' such that $\frac{W_0 g_0}{W_1 g_1} \leq h' < h$, and such that there exists a point a' in some d'-dimensional Euclidean space for which $r_{d'}(a') = h'$, is also in $\bar{K}$. In a similar way it can be shown that, if $h_0 < \frac{W_0 g_0}{W_1 g_1}$ is any point in $\bar{K}$, any point h_0' such that $h_0 < h_0' \leq \frac{W_0 g_0}{W_1 g_1}$, and such that there exists a point a_0' in some d''-dimensional Euclidean space for which $r_{d''}(a_0') = h_0'$, is also in $\bar{K}$. This will prove the lemma.

Let therefore h and h' be as above. Let S^* be the sequential test based on K, with the decision function φ. Let a be a point in d-space such that $r_d(a) = h$. Since $h \in \bar{K}$ we have $\gamma(h) > 0$.

We now wish to define partially another sequential test $\bar{S}$, with a decision function which may be different from φ, as follows: Let a' be defined as above. Write

$$a = (a_1, \cdots, a_d)$$
$$a' = (a_1', \cdots, a_{d'}').$$

Let

$$\bar{a} = a_1, \cdots, a_d, y_1, \cdots, y_t$$

be any sequence such that $n(\bar{a}, S^*) = d + t$. Then for the sequence

$$\bar{a}' = a_1', \cdots, a_{d'}', y_1, \cdots, y_t$$

let $n(\bar{a}', \bar{S}) = d' + t$. The decision function ψ associated with $\bar{S}$ will be partially defined as follows:

$$\psi(\bar{a}') = \varphi(\bar{a}).$$

Clearly

$$E_i(n \mid a, S^*) - d = E_i(n \mid a', \bar{S}) - d' \qquad (i = 0, 1) \tag{4.1}$$

and

$$E_i(\varphi \mid a, S^*) = E_i(\psi \mid a', \bar{S}) \qquad (i = 0, 1). \tag{4.2}$$

Furthermore, we have

$$\begin{aligned} &l(a) - E(L_n \mid a, S^*) \\ &= \frac{g_0}{g_0 + g_1 h} \{W_0 + cd - cE_0(n \mid a, S^*) - W_0[1 - E_0(\varphi \mid a, S^*)]\} \\ &\quad + \frac{g_1 h}{g_0 + g_1 h} \{cd - cE_1(n \mid a, S^*) - W_1 E_1(\varphi \mid a, S^*)\}. \end{aligned} \tag{4.3}$$

Since $\gamma(h) > 0$, and since

$$cd - cE_1(n \mid a, S^*) - W_1 E_1(\varphi \mid a, S^*) < 0, \tag{4.4}$$

we must have

$$W_0 + cd - cE_0(n \mid a, S^*) - W_0[1 - E_0(\varphi \mid a, S^*)] > 0. \tag{4.5}$$

From $h' < h$ it follows that

$$\frac{g_0}{g_0 + g_1 h'} > \frac{g_0}{g_0 + g_1 h}, \qquad \text{and} \qquad \frac{g_1 h'}{g_0 + g_1 h'} < \frac{g_1 h}{g_0 + g_1 h}. \tag{4.6}$$

Relations (4.1), (4.2), (4.4), (4.5) and (4.6) imply that the value of the right hand member of (4.3) is increased by replacing φ, h, a, S^* and d by ψ, h', a', $\bar{S}$, and d', respectively. This proves our lemma.

If there are values which r_j cannot assume the pair B, A might not be unique. For convenience we shall define A and B uniquely in the manner described below. We will always adhere to this definition thereafter.

We shall first define $\gamma(h)$ for all positive h in a manner consistent with the previous definition, which defined $\gamma(h)$ only for those values of h which could be assumed by r_j. Let h be any positive number and $D(h)$ be any sequential test with the following properties:

(4.7) there exists a set $Q(h)$ of positive numbers such that $n = j$ if and only if the j-th member of the sequence

$$hr_1, hr_2, hr_3, \cdots$$

is the first element of the sequence to be in $Q(h)$

$$E_i(n \mid D(h)) < \infty \qquad (i = 0, 1). \tag{4.8}$$

We define, for $h \geq \frac{W_0 g_0}{W_1 g_1}$,

$$\gamma(h \mid D(h)) = \frac{g_0}{g_0 + g_1 h} \{W_0 E_0(\varphi \mid D(h)) - cE_0(n \mid D(h))\} + \frac{g_1 h}{g_0 + g_1 h} \{-W_1 E_1(\varphi \mid D(h)) - cE_1(n \mid D(h))\}, \tag{4.9}$$

$$\gamma(h) = \sup_{D(h)} \gamma(h \mid D(h)) \tag{4.10}$$

with a corresponding definition for $h \leq \frac{W_0 g_0}{W_1 g_1}$. Thus $\gamma(h)$ is defined for all positive h. This definition coincides with the previous definition whenever the latter is applicable. It is true that the supremum operation in (4.10) is limited to tests which depend only on the probability ratio, as (4.7) implies, but the argument of Lemma 1 shows that this limitation does not diminish the supremum. (It might appear that, for $h = \frac{W_0 g_0}{W_1 g_1}$, $\gamma(h)$ is not uniquely defined. We shall shortly see that this is not the case.)

The quantity $\gamma(h)$ depends, of course, on g_0 and g_1. To put this in evidence, we shall also write $\gamma(h, g_0, g_1)$. One can easily verify that

$$\gamma(h, g_0, g_1) = \gamma\left(1, \frac{g_0}{g_0 + g_1 h}, \frac{g_1 h}{g_0 + g_1 h}\right).$$

More generally, for any positive values h and h', we have $\gamma(h, g_0, g_1) = \gamma(h', \bar{g}_0, \bar{g}_1)$, where $\bar{g}_0$ and $\bar{g}_1$ are suitable functions of g_0, g_1, h, and h'. Thus, if h is not an admissible value of the probability ratio and h' is any admissible value, we can interpret the value of $\gamma(h, g_0, g_1)$ as the value of γ corresponding to h' and some properly chosen a priori probabilities $\bar{g}_0$ and $\bar{g}_1$.

We now define A as the greatest lower bound of all points $h \geq \frac{W_0 g_0}{W_1 g_1}$ for which $\gamma(h) \leq 0$. We define B as the least upper bound of all points $h \leq \frac{W_0 g_0}{W_1 g_1}$ for which $\gamma(h) \leq 0$. If $\gamma(h) \leq 0$ for all h the above definition implies $A = B = \frac{W_0 g_0}{W_1 g_1}$.

The argument of Lemma 2 shows that $\gamma(h)$ is monotonically increasing in the interval $\left(B, \frac{W_0 g_0}{W_1 g_1}\right)$ and that $\gamma(h)$ is monotonically decreasing in the interval $\left(\frac{W_0 g_0}{W_1 g_1}, A\right)$.

We shall now define a sequential test $S^*(h)$ for every positive h. The decision

function of $S^*(h)$ will be φ, and $n = j$ if and only if the j-th member of the sequence

$$\gamma(hr_1), \quad \gamma(hr_2), \quad \gamma(hr_3), \cdots$$

is the first element to be ≤ 0. We see that

$$\gamma(h) = \gamma(h \mid S^*(h)) \tag{4.11}$$

for *all h*. Incidentally, this proves that $\gamma(h)$ was uniquely defined at $h = \frac{W_0 g_0}{W_1 g_1}$.

We shall now prove

LEMMA 3. *The function $\gamma(h)$ has the following properties:*

a) It is continuous for all h.

b) $\gamma(A) = \gamma(B) = 0$

c) $\gamma(h) < 0$ for $h > A$ or $< B$.

Only a) and c) require proof, since b) is a trivial consequence of a) and the definition of A and B.

Let h be any point except $\frac{W_0 g_0}{W_1 g_1}$, and let z be any point in a neighborhood of h. Within a neighborhood of h both $E_0(n \mid S^*(z))$ and $E_1(n \mid S^*(z))$ are bounded. Let Δ be an arbitrarily given, positive number. Let h' and h'' be any two points in a sufficiently small neighborhood of h, to be described shortly. We proceed as in the argument of Lemma 2, with the present h' corresponding to h of Lemma 2, the present h'' corresponding to h' of Lemma 2, and with $S^*(h')$ corresponding to S^* of Lemma 2. Since $\frac{g_0}{g_0 + g_1 z}$ and $\frac{g_1 z}{g_0 + g_1 z}$ are continuous functions of z, and since $E_0(n \mid S^*(z))$ and $E_1(n \mid S^*(z))$ are bounded functions of z, we conclude that, when the neighborhood of h is sufficiently small,

$$\gamma(h'') \geq \gamma(h') - \Delta.$$

Reversing the roles of h' and h'' we obtain that in this neighborhood

$$\gamma(h') \geq \gamma(h'') - \Delta,$$

and conclude that

$$|\gamma(h') - \gamma(h'')| \leq \Delta.$$

Since Δ was arbitrary, this implies the continuity of $\gamma(h)$ everywhere, except perhaps at $h = \frac{W_0 g_0}{W_1 g_1}$.

To deal with the point $h = \frac{W_0 g_0}{W_1 g_1}$, proceed as follows: Using the above argument and the definition (4.9), (4.10), we prove that $\gamma(h)$ is continuous on the right

at $h = \frac{W_0 g_0}{W_1 g_1}$. Using, at the point $h = \frac{W_0 g_0}{W_1 g_1}$, the definition of $\gamma(h \mid D(h))$ for $h \leq \frac{W_0 g_0}{W_1 g_1}$ i.e.,

$$\begin{aligned}\gamma(h \mid D(h)) &= \frac{g_0}{g_0 + g_1 h}\{-W_0 E_0(1 - \varphi \mid D(h)) - cE_0(n \mid D(h))\} \\ &\quad + \frac{g_1 h}{g_0 + g_1 h}\{W_1 E_1(1 - \varphi \mid D(h)) - cE_1(n \mid D(h))\},\end{aligned} \tag{4.12}$$

(4.10) and (4.11), we prove that $\gamma(h)$ is continuous on the left at $h = \frac{W_0 g_0}{W_1 g_1}$. This proves a).

To prove c), we proceed as follows: Suppose for $h_0 > A$ we had $\gamma(h_0) = 0$. Since

$$\{-W_1 E_1(\varphi \mid S^*(h_0)) - cE_1(n \mid S^*(h_0))\} < 0,$$

we would have that

$$\{W_0 E_0(\varphi \mid S^*(h_0)) - cE_0(n \mid S^*(h_0))\} > 0.$$

An argument like that of Lemma 2 would then show that $\gamma(h) > 0$ for $\frac{W_0 g_0}{W_1 g_1} < h < h_0$. This, however, is impossible, because it is a violation of the definition of A.

In a similar way we prove that if $h < B$, $\gamma(h) < 0$. This proves c) and with it the lemma.

5. The behavior of A and B. LEMMA 4. *Let g and c be fixed. Then A and B are continuous functions of W_0 and W_1.*

PROOF: It will be sufficient to prove that A is continuous, the proof for B being similar. Suppose $A > B$. Let h_1 and h_2 be such that

a) $B < h_1 < A < h_2$;

b) $h_2 - h_1 < \Delta$ for an arbitrary positive Δ.

We write $\gamma(h)$ temporarily as $\gamma(h, W_0, W_1)$ in order to exhibit the dependence on W_0 and W_1. Then

$$\gamma(h_1, W_0, W_1) > 0;$$
$$\gamma(h_2, W_0, W_1) < 0.$$

It follows from (4.9) that $\gamma(h \mid D(h))$ is continuous in W_0, W_1, uniformly in $D(h)$. Hence $\gamma(h, W_0, W_1) = \sup_{D(h)} \gamma(h \mid D(h))$ is also continuous in W_0, W_1. Hence, for ΔW_0 and ΔW_1 sufficiently small,

$$\gamma(h_1, W_0 + \Delta W_0, W_1 + \Delta W_1) > 0;$$
$$\gamma(h_2, W_0 + \Delta W_0, W_1 + \Delta W_1) < 0.$$

Therefore

$$h_1 \leq A(W_0 + \Delta W_0, W_1 + \Delta W_1) \leq h_2,$$

which proves continuity, since Δ was arbitrary.

If $\frac{W_0 g_0}{W_1 g_1} = A = B$, we take $h_1 < \frac{W_0 g_0}{W_1 g_1} < h_2$, $h_2 - h_1 < \Delta$, and by a similar argument show that

$$\gamma(h_1, W_0 + \Delta W_0, W_1 + \Delta W_1) < 0;$$
$$\gamma(h_2, W_0 + \Delta W_0, W_1 + \Delta W_1) < 0.$$

Thus

$$h_1 \leq B(W_0 + \Delta W_0, W_1 + \Delta W_1) \leq A(W_0 + \Delta W_0, W_1 + \Delta W_1) \leq h_2.$$

This proves the lemma.

LEMMA 5. *Let g, c, and W_1 be fixed. A is strictly monotonic in W_0. As W_0 approaches 0, A approaches 0; as W_0 approaches $+\infty$, A also approaches $+\infty$.*

PROOF: Since $A \geq \frac{W_0 g_0}{W_1 g_1}$, $A \to +\infty$ as $W_0 \to +\infty$. If $W_0 < c$ no reduction in average risk could compensate for taking even a single observation, no matter what the value of h. Hence $\gamma(h) \leq 0$ for all h when $W_0 < c$, so that $A = B$. Since $B \leq \frac{W_0 g_0}{W_1 g_1}$, $B \to 0$ as $W_0 \to 0$. Hence $A \to 0$ as $W_0 \to 0$. It is evident from (4.9) that $\gamma(h \mid D(h))$ is non-decreasing with increasing W_0 (everything else fixed). Hence also

$$\gamma(h) = \sup_{D(h)} \gamma(h \mid D(h)),$$

is non-decreasing with increasing W_0, for fixed $h > \frac{W_0 g_0}{W_1 g_1}$ and fixed W_1. For a positive Δ sufficiently small and for any h such that $A \leq h < A + \Delta$, we have that

$$E_0(\varphi \mid S^*(h)) > 0.$$

Hence, for such h, $\gamma(h, W_0, W_1)$ is strictly monotonically increasing with increasing W_0. Therefore A is (strictly) monotonically increasing with increasing W_0.

We now define the function $W_0(W_1, \delta)$ of the two positive arguments W_1, δ so that

$$A(W_0(W_1, \delta), W_1) = \delta.$$

By Lemma 5 such a function exists and is single-valued.

6. Properties of the function $W_0(W_1, \delta)$. LEMMA 6. *$W_0(W_1, \delta)$ is continuous in W_1.*

PROOF: Let

$$\lim_{N \to \infty} W_{1N} = W_1,$$

and suppose that the sequence $\{W_0(W_{1N}, \delta)\}$ did not converge. Suppose W_0' and W_0'' were two distinct limit points of this sequence. From the continuity of A (Lemma 4) it would follow that

$$A(W_0', W_1) = A(W_0'', W_1)$$

This, however, violates Lemma 5. The only remaining possibility to be considered is that

$$\lim_{N \to \infty} W_0(W_{1N}, \delta) = \infty.$$

If that were the case, then, since $A \geq \dfrac{W_0 g_0}{W_1 g_1}$, it would follow that $A \to \infty$, in violation of the fact that $A \equiv \delta$.

Lemma 7. *We have, for fixed δ,*

$$\lim_{W_1 \to 0} W_0(W_1) = 0;$$

$$\lim_{W_1 \to \infty} W_0(W_1) = \infty.$$

Proof: If, for small W_1, $W_0(W_1)$ were bounded below by a positive number, then, since $A \geq \dfrac{g_0 W_0(W_1, \delta)}{W_1 g_1}$, we could make A arbitrarily large by taking W_1 sufficiently small, in violation of the fact that $A \equiv \delta$. To prove the second half of the lemma, assume that $W_0(W_1)$ is bounded above as $W_1 \to \infty$. Then $B\left(\leq \dfrac{W_0 g_0}{W_1 g_1}\right)$ will approach zero as $W_1 \to \infty$. Let h be fixed so that $B < h < \delta$. Consider the totality of points ω for which there exists an integer $n^*(\omega)$ such that:

$$hr_{n^*} \leq B;$$

$$B < hr_j < \delta, \qquad j < n^*.$$

The conditional expected value of n^* in this totality, when H_0 is true, may be made arbitrarily large by making B sufficiently small. Hence, when W_1 is sufficiently large, for fixed but arbitrary $h < \delta$, the optimum procedure from the point of minimizing the average risk is to reject H_0 at once without taking any more observations. This, however, contradicts the fact that $h < \delta$, and proves the lemma.

Lemma 8. *We have, for fixed $\delta > 1$,*

$$\lim_{W_1 \to 0} B(W_0(W_1, \delta), W_1) = \delta;$$

$$\lim_{W_1 \to \infty} B(W_0(W_1, \delta), W_1) = 0.$$

Proof: By Lemma 7,

$$\lim_{W_1 \to 0} W_0(W_1) = 0.$$

When, for fixed c, both W_0 and W_1 are small enough, then, no matter what the value of h, $\gamma(h) < 0$. Hence $A = B$, which proves the first half of the lemma.

Let now $\{W_{1N}\}$ be a sequence such that $\lim W_{1N} = \infty$. Let $\delta > 1$. For the sake of brevity we write $B(W_{1N})$ instead of

$$B(W_0(W_{1N}\delta), W_{1N}).$$

Suppose that, for sufficiently large N, $B(W_{1N})$ is bounded below by a positive number. Hence, for sufficiently large N, the probability of rejecting H_1 when it it is true is bounded below by a positive number. Moreover, since $B \leq \frac{W_0 g_0}{W_1 g_1} \leq A$, it follows that, for N sufficiently large, $\frac{W_{0N} g_0}{W_{1N} g_1}$ is bounded above and below by positive constants. Thus, for large N the average risk of the test defined by $B(W_{1N})$, δ, is greater than ug_1W_{1N}, where u is a positive constant which does not depend on N. Moreover, from the definition of $B(W_{1N})$, this risk is a minimum.

Let ϵ be a positive number such that $\epsilon\left(\frac{W_{0N}g_0}{W_{1N}g_1} + 1\right) < \frac{u}{2}$ for all N sufficiently large. Let V_1, V_2, with $0 < V_1 < 1 < V_2$, be two constants such that, for the sequential probability ratio test determined by them, both α_0 and α_1 are $< \epsilon$. Of course E_0n and E_1n are finite and determined by the test. For this test the average risk is less than

$$\begin{aligned}\epsilon(g_0 W_{0N} + g_1 W_{1N}) &+ cg_0 E_0 n + cg_1 E_1 n \\ &< \frac{u}{2} g_1 W_{1N} + cg_0 E_0 n + cg_1 E_1 n \\ &< \frac{3u}{4} g_1 W_{1N},\end{aligned}$$

for W_{1N} large enough. This however contradicts the fact that the minimum risk is $> ug_1W_{1N}$, and proves the lemma.

7. Proof of the theorem. Let a given sequential probability ratio test S_0 be defined by B^*, A^*; $B^* < 1 < A^*$. Let $\alpha_i(S_0)$ be the probability, according to S_0, of rejecting H_i when it is true. Let c be fixed.

By Lemma 4, B is a continuous function of W_0 and W_1. Let $\delta = A^*$ in Lemma 8. Then there exists a pair $\overline{W}_0$, $\overline{W}_1$, with $\overline{W}_0 = W_0(\overline{W}_1, A^*)$, such that

$$A(\overline{W}_0, \overline{W}_1) = A^*;$$

$$B(\overline{W}_0, \overline{W}_1) = B^*.$$

Hence the average risk

$$\sum_i g_i[\overline{W}_i \alpha_i(S_0) + cE_i^0(n)],$$

corresponding to the sequential test S_0 is a minimum.

Now let S_1 be any other test for deciding between H_0 and H_1 and such that

$$\alpha_i(S_1) \leq \alpha_i(S_0), \quad \text{and} \quad E_i^1(n) \text{ exists } (i = 1, 2).$$

Then

$$\sum_i g_i [\overline{W}_i \, \alpha_i(S_0) + cE_i^0(n)] \leq \sum_i g_i [\overline{W}_i \, \alpha_i(S_1) + cE_i^1(n)].$$

Since $\alpha_i(S_1) \leq \alpha_i(S_0)$, we have

$$\sum_i g_i \, E_i^0(n) \leq \sum_i g_i \, E_i^1(n).$$

Now g_0, g_1 were arbitrarily chosen (subject, of course, to the obvious restrictions). Hence it must be that

$$E_i^0(n) \leq E_i^1(n).$$

This, however, is the desired result.

REFERENCES

[1] A. WALD, "Sequential tests of statistical hypotheses", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 117–186.

[2] A. WALD, *Sequential Analysis*, John Wiley and Sons, Inc., New York, 1947.

[3] CHARLES STEIN, "A note on cumulative sums", *Annals of Math. Stat.*, Vol. 17 (1946), pp. 498–499.

LIMITING DISTRIBUTION OF A ROOT OF A DETERMINANTAL EQUATION

By D. N. Nanda

Institute of Statistics, University of North Carolina

1. Summary. The exact distribution of a root of a determinantal equation when the roots are arranged in a monotonic order was obtained by S. N. Roy [3] in 1943. A different method for deriving the distribution of any one of these roots has been described by the author in [2]. In the present paper the limiting forms of these distributions are obtained. This paper gives a method by which the limiting distributions can be obtained without undergoing an inordinate amount of mathematical labor.

2. Introduction. If $x = \| x_{ij} \|$ and $x^* = \| x^*_{ij} \|$ are two p-variate sample matrices with n_1 and n_2 degrees of freedom and $S = \| xx' \|/n_1$ and $S^* = \|x^*x^{*\prime}\|/n_2$ are the covariance matrices which under the null hypothesis are independent estimates of the same population covariance matrix, then the joint distribution of the roots of the determinantal equation $| A - \theta(A + B) | = 0$, where $A = n_1S$ and $B = n_2S^*$, was obtained by Hsu [1] in 1939 and is

$$(1) \quad R'(l, \mu, \nu) = \frac{\pi^{l/2} \prod_{i=1}^{l} \Gamma\left(\frac{l+\mu+\nu+i-2}{2}\right)}{\prod_{i=1}^{l} \Gamma\left(\frac{\mu+i-1}{2}\right) \Gamma\left(\frac{\nu+i-1}{2}\right) \Gamma\left(\frac{i}{2}\right)}$$

$$\prod_{i=1}^{l} (\theta_i)^{\mu/2-1} \prod_{i=1}^{l} (1-\theta_i)^{\nu/2-1} \prod_{i<j} (\theta_i - \theta_j),$$

$$(0 \leq \theta_l \leq \theta_{l-1} \leq \cdots \leq \theta_1 \leq 1),$$

where $l = \min(p, n_1)$, $\mu = | p - n_1 | + 1$ and $\nu = n_2 - p + 1$. The distribution density may be expressed as

$$(2) \qquad R(l, m, n) = c(l, m, n) \prod_{i=1}^{l} [\theta_i^m (1 - \theta_i)^n] \prod_{i<j} (\theta_i - \theta_j),$$

where $m = \mu/2 - 1$ and $n = \nu/2 - 1$.

3. Method. Let $\theta_i = \zeta_i/n$ in (2). The joint distribution reduces to

$$(3) \qquad \frac{c(l, m, n)}{n^{l+lm+l(l-1)/2}} \prod_{i=1}^{l} [\zeta_i^m (1 - \zeta_i/n)^n] \prod_{i<j} (\zeta_i - \zeta_j)\, d\zeta_1 \cdots d\zeta_l,$$

$$(0 \leq \zeta_l \leq \zeta_{l-1} \cdots \leq \zeta_1 \leq n).$$

As n tends to infinity the limit of (3) is

$$K(l, m) \quad \prod_{i=1}^{l} \zeta_i^m \prod_{i<j} (\zeta_i - \zeta_j) e^{-\Sigma \zeta_i} \, d\zeta_i \,. \tag{4}$$

$$(0 \leq \zeta_l \leq \zeta_{l-1} \cdots \leq \zeta_1 < \infty).$$

The value of $K(l, m)$ is

$$\lim_{n\to\infty} \frac{c(l, m, n)}{n^{l+lm+l(l-1)/2}}$$

$$= \lim_{n\to\infty} \frac{\pi^{l/2} \prod_{i=1}^{l} \Gamma\left(\frac{l+2m+2n+i+2}{2}\right)}{\prod_{i=1}^{l} \Gamma\left(\frac{2m+i+1}{2}\right) \Gamma\left(\frac{2n+i+1}{2}\right) \Gamma(i/2) \cdot n^{l+lm+l(l-1)/2}}$$

$$= \frac{\pi^{l/2}}{\prod_{i=1}^{l} \Gamma\left(\frac{2m+i+1}{2}\right) \Gamma(i/2)} \cdot \lim_{n\to\infty} \frac{\prod_{i=1}^{l} \Gamma\left(\frac{l+2m+2n+i+2}{2}\right)}{\prod_{i=1}^{l} \Gamma\left(\frac{2n+i+1}{2}\right) \cdot n^{l+lm+l(l-1)/2}}$$

By using Stirling's approximation for gamma functions and after simplification we get

$$\lim_{n\to\infty} \frac{\prod_{i=1}^{l} \Gamma\left(\frac{l+2m+2n+i+2}{2}\right)}{\prod_{i=1}^{l} \Gamma\left(\frac{2n+i+1}{2}\right) \cdot n^{lm+l(l+1)/2}} = 1.$$

Hence

$$K(l, m) = \frac{\pi^{l/2}}{\prod_{i=1}^{l} \Gamma\left(\frac{2m+i+1}{2}\right) \Gamma(i/2)},$$

and therefore

$$\begin{aligned} K(2, m) &= 2^{2m+1}/\Gamma(2m+2), \\ K(3, m) &= 2^{2m+3}/[\Gamma(m+1)\Gamma(2m+3)], \\ K(4, m) &= 2^{4m+5}/[\Gamma(2m+2)\Gamma(2m+4)], \\ K(5, m) &= 2^{4m+9}/[3\Gamma(m+1)\Gamma(2m+3)\Gamma(2m+5)]. \end{aligned} \tag{5}$$

Let

$$G_{l,m}(x) = K(l, m) \int_{0 \leq \zeta_l \leq \zeta_{l-1} \cdots < \zeta_1 \leq x} \prod_{i=1}^{l} \zeta_i^m \prod_{i<j} (\zeta_i - \zeta_j) e^{-\Sigma \zeta_i} \prod d\zeta_i \,. \tag{6}$$

It can easily be observed that

$$G_{l,m}(x) = \Pr(\zeta_1 \leq x) = \lim_{n\to\infty} \Pr(n\theta_1 \leq x) = \lim_{n\to\infty} \Pr\left(\theta_1 \leq \frac{x}{n}\right).$$

Thus the limiting form of the distribution of the largest root can be obtained by integrating the density given in (4) according to the method described by the author in [2]. It is, however, observed that the mathematical labor is reduced considerably by adopting the following method.

Referring to the results of the exact distribution of the largest root given in [2], let $F_{l,m,n}(x) = (0, l, l-1, \cdots, 1, x; m, n)$; thus $F_{2,m,n}(x) = (0, 2, 1, x; m, n)$ and $F_{3,m,n}(x) = (0, 3, 2, 1, x; m, n)$. Then $c(l, m, n)F_{l,m,n}(x)$ is the probability that none of the roots θ_i exceeds x, and is thus the cumulative distribution function of the greatest root. We shall show that $\lim_{n\to\infty} c(l, m, n)F_{l,m,n}(x/n) = G_{l,m}(x)$. The reader is, however, asked to refer to [2] for the detailed explanation of the notations and certain mathematical operations used in this paper.

4. Limiting distribution of the largest root. We will derive the distribution of the largest root for $l = 2$ and 3 by the two methods. A straightforward method will be named A. A second method, which proves to be very simple and easy will be called Method B.

(a) $l = 2$

(i) Method A. We have,

$$\Pr(n\theta_1 \leq x) = G_{2,m}(x) = K(2, m) \int_{0<\zeta_2<\zeta_1<x} (\zeta_1\zeta_2)^m(\zeta_1 - \zeta_2)e^{-(\zeta_1+\zeta_2)}\, d\zeta_1\, d\zeta_2$$

By using the method described in [2], we have

$$\begin{aligned} G_{2,m}(x) &= K(2, m)\left\{\int_{0<\zeta_2<\zeta_1<x} - \int_{0<\zeta_1<\zeta_2<x} \zeta_2^m e^{-\zeta_2}\cdot\zeta_1^{m+1}e^{-\zeta_1}\, d\zeta_1\, d\zeta_2\right\}, \\ &= K(2, m)\{T_0^{m,x}(y, 1, x; m+1) - T_0^{m,x}(0, 1, y; m+1)\}, \end{aligned}$$

where

$$T_a^{m,b}g(y) = \int_a^b g(y)\cdot y^m e^{-y}\, dy,$$

and

$$(7)\,(a, 1, b; m+1) = \int_a^b \zeta^{m+1}e^{-\zeta}\, d\zeta = (a^{m+1}e^{-a} - b^{m+1}e^{-b}) + (m+1)(a, 1, b; m).$$

Hence,

$$\begin{aligned} G_{2,m}(x) &= K(2, m)T_0^{m,x}[y^{m+1}e^{-y} - x^{m+1}e^{-x} + (m+1)(y, 1, x; m) + y^{m+1}e^{-y} \\ &\qquad - (m+1)(0, 1, y; m)], \\ &= K(2, m)T_0^{m,x}[2y^{m+1}e^{-y} - x^{m+1}e^{-x}], \end{aligned}$$

as $T_0^{m,x}[(y, 1, x; m) - (0, 1, y; m)] = 0$.

Therefore

$$\lim_{n\to\infty} \Pr(n\theta_1 \le x) = G_{2,m}(x) = K(2, m) \tag{8}$$
$$\cdot \left\{ 2\int_0^x y^{2m+1}e^{-2y}dy - x^{m+1}e^{-x} \int_0^x y^m e^{-y}\, dy \right\}.$$

When $x = \infty$, $G_{2,m}(x) = 1$; hence $K(2, m) = 2^{2m+1}/\Gamma(2m + 2)$, the value given in (5).

Now we shall derive the result by Method B.

(ii) METHOD B.

$$F_{2,m,n}(x) = (0, 2, 1, x; m, n) = \frac{1}{m + n + 2}$$
$$\cdot \left\{ 2 \int_0^x y^{2m+1}(1 - y)^{2n+1}\, dy - x^{m+1}(1 - x)^{n+1} \int_0^x y^m(1 - y)^n\, dy \right\},$$

a result given in [2].

Replacing x by x/n, we get

$$(0, 2, 1, x/n; m, n) = \frac{1}{m + n + 2}$$
$$\cdot \left\{ 2 \int_0^{x/n} y^{2m+1}(1 - y)^{2n+1}\, dy - (x/n)^{m+1}(1 - x/n)^{n+1} \int_0^{x/n} y^m(1 - y)^n\, dy \right\};$$

also, letting $y = u/n$, we have

$$(0, 2, 1, x/n; m, n) = \frac{1}{(m + n + 2)n^{2m+2}} \tag{9}$$
$$\cdot \left\{ 2 \int_0^x u^{2m+1}(1 - u/n)^{2n+1}\, du - x^{m+1}(1 - x/n)^{n+1} \int_0^x u^m(1 - u/n)^n\, du \right\}.$$

Thus

$$\lim_{n\to\infty} \Pr(n\theta_1 \le x) = \Pr(\theta_1 \le x/n) = \lim_{n\to\infty} c(2, m, n)(0, 2, 1, x/n; m, n),$$
$$= \frac{2^{2m+1}}{\Gamma(2m + 2)} \left\{ 2 \int_0^x u^{2m+1}e^{-2u}\, du - x^{m+1}e^{-x} \int_0^x u^m e^{-u}\, du \right\},$$

which is the same as (8), obtained by Method A.

(b) $l = 3$.

(i) METHOD A. We have

$$\Pr(n\theta_1 \le x) = G_{3,m}(x) = K(3, m) \int_{0<\zeta_3<\zeta_2<\zeta_1<x} \Pi\zeta_i^m \Pi(\zeta_i - \zeta_j)e^{-\Sigma\zeta_i}\Pi\, d\zeta_i$$
$$= K(3, m) \int_{0<\zeta_3<\zeta_2<\zeta_1<x} (\zeta_1\zeta_2\zeta_3)^m e^{-(\zeta_1+\zeta_2+\zeta_3)} \{1, 2, 3\}\, d\zeta_1\, d\zeta_2\, d\zeta_3,$$

where $\{1, 2, 3\} = \zeta_1\zeta_2\{1, 2\} + \zeta_3\zeta_1\{3, 1\} + \zeta_2\zeta_3\{2, 3\}$, as given in [2].

Or

$$G_{3,m(x)} = K(3, m)\left\{\int_{0<\zeta_3<\zeta_2<\zeta_1<x} + \int_{0<\zeta_1<\zeta_3<\zeta_2<x} + \int_{0<\zeta_2<\zeta_1<\zeta_3<x} \zeta_3^m e^{-\zeta_3}(\zeta_1\zeta_2)^{m+1}e^{-(\zeta_1+\zeta_2)}\{1, 2\}\, d\zeta_1\, d\zeta_2\, d\zeta_3\right\},$$

$$= K(3, m)\{T_0^{m,x}(y, 2, 1, x; m + 1) + T_0^{m,x}(0, 2, y, 1, x; m + 1) + T_0^{m,x}(0, 2, 1, y; m + 1)\},$$

where

$$(a, 2, 1, b; m) = \int_{a<\zeta_2<\zeta_1<b} (\zeta_1\zeta_2)^m(\zeta_1-\zeta_2)e^{-(\zeta_1+\zeta_2)}\, d\zeta_1\, d\zeta_2.$$

We have already obtained

$$(0, 2, 1, x; m) = G_{2,m}(x)/K(2, m) = \left\{2\int_0^x y^{2m+1}e^{-2y}\, dy - x^{m+1}e^{-x}\int_0^x y^m e^{-y}\, dy\right\}$$

as given in (8).

We also need the following results which are obtained by the method described for $l = 2$.

$$(10)\qquad (a, 2, 1, b; m) = \left\{2\int_a^b u^{2m+1}e^{-2u}\, du - (a^{m+1}e^{-a} + b^{m+1}e^{-b})\int_a^b u^m e^{-u}\, du\right\},$$

and

$$(11)\qquad (a, 2, b, 1, c; m) = \left\{b^{m+1}e^{-b}\int_a^c u^m e^{-u}\, du - a^{m+1}e^{-a}\int_b^c u^m e^{-u}\, du - c^{m+1}e^{-c}\int_a^b u^m e^{-u}\, du\right\}.$$

Using these results we have

$$G_{3,m}(x) = K(3, m)T_0^{m,x}\left\{2\int_y^x u^{2m+3}e^{-2u}\, du - (y^{m+2}e^{-y} + x^{m+2}e^{-x})\int_y^x u^{m+1}e^{-u}\, du - y^{m+2}e^{-y}\int_0^x u^{m+1}e^{-u}\, du + x^{m+2}e^{-x}\int_0^y u^{m+1}e^{-u}\, du + 2\int_0^y u^{2m+3}e^{-2u}\, du - y^{m+2}e^{-y}\int_0^y u^{m+1}e^{-u}\, du\right\}.$$

Simplifying we get

$$(12)\quad \lim_{n\to\infty} \Pr(n\theta_1 \le x) = G_{3,m}(x) = K(3, m)\left\{2\int_0^x u^{2m+3}e^{-2u}\, du\int_0^x u^m e^{-u}\, du - 2\int_0^x u^{m+1}e^{-u}\, du\int_0^x u^{2m+2}e^{-2u}\, du - x^{m+2}e^{-x}\left[2\int_0^x u^{2m+1}e^{-2u}\, du - x^{m+1}e^{-x}\int_0^x u^m e^{-u}\, du\right]\right\},$$

where $K(3, m) = 2^{2m+3}/[\Gamma(2m + 1)\Gamma(2m + 3)]$.

(ii) METHOD B.

$$
\begin{aligned}
F_{3,m,n}(x) &= (0, 3, 2, 1, x; m, n) \\
&= \frac{1}{m+n+3}\,[2(0, 1, x; 2m+3, 2n+1)(0, 1, x; m, n) \\
&\quad -2(0, 1, x; m+1, n)(0, 1, x; 2m+2, 2n+1) \\
&\qquad - (0, x; m+2, n+1)(0, 2, 1, x; m, n)],
\end{aligned}
$$

a result given in [2].

Replacing x by x/n and putting u/n for the variate y of integration, we have,

$$
\begin{aligned}
F_{3,m,n}(x) &= (0, 3, 2, 1; x/n; m, n) = \frac{1}{m+n+3} \\
&\left\{\frac{2}{n^{3m+5}} \int_0^x u^{2m+3}(1-u/n)^{2n+1}\,du \int_0^x u^m(1-u/n)^n\,du - \frac{2}{n^{3m+5}}\right. \\
&\int_0^x u^{m+1}(1-u/n)^n\,du \int_0^x u^{2m+2}(1-u/n)^{2n+1}\,du - \frac{x^{m+2}(1-x/n)^{n+1}}{n^{3m+4}(m+n+2)} \\
&\left.\left[2\int_0^x u^{2m+1}(1-u/n)^{2n+1}\,du - x^{m+1}(1-x/n)^{n+1}\int_0^x u^m(1-u/n)^n\,du\right]\right\}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
\lim_{n\to\infty} \Pr\,(n\theta_1 \le x) &= \lim_{n\to\infty} \Pr\left(\theta_1 \le \frac{x}{n}\right) = \lim_{n\to\infty} c(3, m, n)F_{3,m,n}(x) \\
&= K(3, m)\left\{2\int_0^x u^{2m+3}e^{-2u}\,du \int_0^x u^m e^{-u}\,du - 2\int_0^x u^{2m+2}e^{-2u}\,du \int_0^x u^{m+1}e^{+u}\,du\right. \\
&\qquad \left. - x^{m+2}e^{-x}\left[2\int_0^x u^{2m+1}e^{-2u}\,du - x^{m+1}e^{-x}\int_0^x u^m e^{-u}\,du\right]\right\},
\end{aligned}
$$

where

$$K(3, m) = 2^{2m+3}/[\Gamma(m+1)\Gamma(2m+3)].$$

This result is the same as (12) obtained by Method A.

We have thus shown that Method B is applicable for obtaining the limiting forms of the distribution of the largest root and that it is much simpler as compared to the straightforward method called Method A here.

The limiting distributions for the largest root for $l = 4$ and 5 are listed below.

(c) $l = 4$.

$$
\begin{aligned}
\lim_{n\to\infty} \Pr\,(n\theta_1 \le x) &= \lim_{n\to\infty} \Pr\left(\theta_1 \le \frac{x}{n}\right) = G_{4,m}(x) \\
&= K(4, m)\left\{2\int_0^x u^{2m+5}e^{-2u}\,du\,\frac{G_{2,m}(x)}{K(2, m)} - 2\int_0^x u^{2m+4}e^{-2u}\,du\right. \\
&\left[2\int_0^x u^{2m+2}e^{-2u}\,du - x^{m+2}e^{-x}\int_0^x u^m e^{-u}\,du + (m+2)\frac{G_{2,m}(x)}{K(2, m)}\right] \\
&\qquad \left. + 2\int_0^x u^{2m+3}e^{-2u}\,du\,\frac{G_{2,m+1}(x)}{K(2, m+1)} - x^{m+3}e^{-x}\,\frac{G_{3,m}(x)}{K(3, m)}\right\},
\end{aligned}
$$

where

$$K(4, m) = 2^{4m+5}/[\Gamma(2m + 2)\Gamma(2m + 4)].$$

(d) $l = 5$.

$$\lim_{n\to\infty} \Pr (n\theta_1 \le x) = \lim_{n\to\infty} \Pr \left(\theta_1 \le \frac{x}{n}\right) = G_{5,m}(x)$$

$$= K(5, m) \left\{2 \int_0^x u^{2m+7} e^{-2u}\, du \frac{G_{3,m}(x)}{k(3, m)} - 2 \int_0^x u^{2m+6} e^{-2u}\, du \right.$$

$$\cdot\left[2 \int_0^x u^{2m+4} e^{-2u}\, du \int_0^x u^m e^{-u}\, du - 2 \int_0^x u^{2m+3} e^{-2u}\, du \right.$$

$$\left. \cdot \int_0^x u^{m+1} e^{-u}\, du - x^{m+3} e^{-x} \frac{G_{2,m}(x)}{K(2, m)} + (m + 3) \frac{G_{3,m}(x)}{K(3, m)}\right]$$

$$+ 2 \int_0^x u^{2m+5} e^{-2u}\, du \left\{2 \int_0^x u^{2m+5} e^{-2u}\, du \int_0^x u^m e^{-u}\, du \right.$$

$$- 2 \int_0^x u^{2m+3} e^{-2u}\, du \int_0^x u^{m+2} e^{-u}\, du - x^{m+3} e^{-x}$$

$$\left.\cdot\left[2 \int_0^x u^{2m+2} e^{-2u}\, du - x^{m+2} e^{-x} \int_0^x u^m e^{-u}\, du + (m + 2) \frac{G_{2,m}(x)}{K(2, m)}\right]\right\}$$

$$\left. - 2 \frac{G_{3,m+1}(x)}{K(3, m + 1)} \int_0^x u^{2m+4} e^{-2u}\, du - x^{m+4} e^{-x} \frac{G_{4,m}(x)}{K(4, m)}\right\},$$

where

$$K(5, m) = 2^{4m+9}/[3\Gamma(m + 1)\Gamma(2m + 3)\Gamma(2m + 5)].$$

5. Limiting distribution of the smallest root. It was shown in [2] that the exact distribution of the smallest root can be obtained by using the relation

$$\Pr (\theta_l \le x) = 1 - \Pr (\theta_1 \le 1 - x \mid \nu, \mu).$$

This relation, however, does not help in obtaining the limiting distribution of the smallest root from that of the largest root. The limiting distribution of the smallest root can be obtained by the method illustrated below.

(a) $l = 2$.

The exact distribution of the smallest root θ_2 can be expressed as

$$\Pr (\theta_2 \le x) = c(2, m, n)\{(0, 2, 1, x; m, n) + (0, 2, x, 1, z; m, n)\},$$

where $z = 1$. Replacing x by x/n, we get

$$\Pr (\theta_2 \le x/n) = c(2, m, n)\{(0, 2, 1, x/n; m, n) + (0, 2, x/n, 1, z; m, n)\},$$

where

$$(0, 2, 1, x/n; m, n) = \frac{1}{m + n + 2}\left[2 \int_0^{x/n} y^{2m+1}(1 - y)^{2n+1}\, dy \right.$$

$$\left. - (0, x/n; m + 1, n + 1) \int_0^{x/n} y^m(1 - y)^n\, dy\right],$$

and

$$(0, 2, x/n, 1, z; m, n) = \frac{1}{m+n+2}\Big[(0, x/n; m+1, n+1)$$

$$\cdot \int_0^z y^m(1-y)^n\,dy - (0, z; m+1, n+1)\int_0^{x/n} y^m(1-y)^n\,dy\Big],$$

as obtained from (6) of [2].

The limiting distribution of θ_2 is

$$\Pr(\theta_2 \le x/n) = \lim_{n\to\infty} c(2, m, n)\{(0, 2, 1, x/n; m, n) + (0, 2, x/n, 1, z; m, n)\}. \tag{13}$$

Putting u/n for y, the variate of integration and allowing n to tend to infinity, we have

$$\lim_{n\to\infty} c(2, m, n)(0, 2, 1, x/n; m, n)$$

$$= K(2, m)\left\{2\int_0^x u^{2m+1}e^{-2u}\,du - x^{m+1}e^{-x}\int_0^x u^m e^{-u}\,du\right\},$$

and

$$\lim_{n\to\infty} c(2, m, n)(0, 2, x/n, 1, z; m, n) = K(2, m)x^{m+1}e^{-x}\int_0^\infty u^m e^{-u}\,du$$

$$= K(2, m)x^{m+1}e^{-x}\,\Gamma(m+1).$$

Substituting these results in (13) we have

$$\lim_{n\to\infty} \Pr(n\theta_2 \le x) = \lim_{n\to\infty} \Pr(\theta_2 \le x/n)$$

$$= K(2, m)\left\{2\int_0^x u^{2m+1}e^{-2u}\,du - x^{m+1}e^{-x}\int_0^x u^m e^{-u}\,du + x^{m+1}e^{-x}\,\Gamma(m+1)\right\},$$

where

$$K(2, m) = 2^{2m+1}/[\Gamma(2m+2)].$$

(b) $l = 3$.

The exact distribution of the smallest root can be expressed as

$$\Pr(\theta_3 \le x) = c(3, m, n)[(0, 3, 2, 1, x; m, n) + (0, 3, 2, x, 1, z; m, n) + (0, 3, x, 2, 1, z; m, n)],$$

where $z = 1$.

Replacing x by x/n and allowing n to tend to infinity we have

$$\Pr(n\theta_3 \le x) = \lim_{n\to\infty} c(3, m, n)[(0, 3, 2, 1, x/n; m, n) + (0, 3, 2, x/n, 1, z; m, n) + (0, 3, x/n, 2, 1, z; m, n)]. \tag{14}$$

The values of these components on the right hand side of the above equation are given below.

$$\lim_{n\to\infty} c(3, m, n)(0, 3, 2, 1, x/n; m, n) = G_{3,m}(x), \quad \text{given by} \quad (12),$$

$$\lim_{n\to\infty} c(3, m, n)(0, 3, 2, x/n, 1, z; m, n)$$

$$\begin{aligned} &= K(3, m)\Big\{\int_x^\infty u^m e^{-u}\, du \Big[2\int_0^x u^{2m+3} e^{-2u}\, du \\ &- x^{m+2} e^{-x}\int_0^x u^{m+1} e^{-u}\, du\Big] - x^{m+2} e^{-x}\Big[2\int_0^x u^{2m+1} e^{-2u}\, du \\ &- x^{m+1} e^{-x}\int_0^x u^m e^{-u}\, du\Big] + x^{m+2} e^{-x}\int_x^\infty u^{m+1} e^{-u}\, du \int_0^x u^m e^{-u}\, du \\ &- 2\int_x^\infty u^{m+1} e^{-u}\, du \int_0^x u^{2m+2} e^{-2u}\, du\Big\}, \end{aligned} \tag{15}$$

and

$$\begin{aligned} &\lim_{n\to\infty} c(3, m, n)(0, 3, x/n, 2, 1, z; m, n) = K(3, m)\Big\{\int_0^x u^m e^{-u}\, du\Big[2\int_x^\infty u^{2m+3} e^{-2u}\, du \\ &- x^{m+2} e^{-x}\int_x^\infty u^{m+1} e^{-u}\, du\Big] - x^{m+2} e^{-x}\Big[2\int_x^\infty u^{2m+1} e^{-2u}\, du - x^{m+1} e^{-x}\int_x^\infty u^m e^{-u}\, du\Big] \\ &+ x^{m+2} e^{-x}\int_0^x u^{m+1} e^{-u}\, du \int_x^\infty u^m e^{-u}\, du - 2\int_0^x u^{m+1} e^{-u}\, du \int_x^\infty u^{2m+2} e^{-2u}\, du\Big\}. \end{aligned}$$

Substituting in (14) we have,

$$\begin{aligned} &\lim_{n\to\infty} \Pr(n\theta_3 \le x) = \{2^{2m+3}/[\Gamma(m+1)\Gamma(2m+3)]\} \\ &\cdot\Big\{2\int_0^\infty u^{2m+3} e^{-2u}\, du \int_0^x u^m e^{-u}\, du + 2\int_0^x u^{2m+3} e^{-2u}\, du \int_x^\infty u^m e^{-u}\, du \\ &- 2\int_0^\infty u^{m+1} e^{-u}\, du \int_0^x u^{2m+2} e^{-2u}\, du - 2\int_0^x u^{m+1} e^{-u}\, du \int_x^\infty u^{2m+2} e^{-2u}\, du \\ &- 2x^{m+2} e^{-x}\int_0^\infty u^{2m+1} e^{-2u}\, du - 2x^{m+2} e^{-x}\int_0^x u^{2m+1} e^{-2u}\, du \\ &+ x^{2m+3} e^{-2x}\Big(\int_0^x u^m e^{-u}\, du + \int_0^\infty u^m e^{-u}\, du\Big)\Big\}. \end{aligned}$$

Or,

$$\begin{aligned} &\lim_{n\to\infty} \Pr(n\theta_3 \le x) = 2^{2m+3}/[\Gamma(m+1)\Gamma(2m+3)] \\ &\cdot\Big\{\frac{\Gamma(2m+4)}{2^{2m+4}}\int_0^x u^m e^{-u}\, du + 2\int_0^x u^{2m+3} e^{-2u}\, du \int_x^\infty u^m e^{-u}\, du \\ &- 2\Gamma(m+2)\int_0^x u^{2m+2} e^{-2u}\, du - 2\int_0^x u^{m+1} e^{-u}\, du \int_x^\infty u^{2m+2} e^{-2u}\, du \\ &- \frac{\Gamma(2m+2)}{2^{2m+1}} x^{m+2} e^{-x} - x^{m+2} e^{-x}\int_0^x u^{2m+1} e^{-2u}\, du + \Gamma(m+1)x^{2m+3} e^{-2x} \\ &+ x^{2m+3} e^{-2x}\int_0^x u^m e^{-u}\, du\Big\}. \end{aligned}$$

Thus we have seen that this method can be used for obtaining the limiting distribution of the smallest root for any value of l.

6. Limiting distribution of any intermediate root. The above method can also be used for obtaining the limiting distribution of any intermediate root. We shall give the distribution of θ_2 for $l = 3$. We have

$$\Pr(\theta_2 \le x) = c(3, m, n)\{(0, 3, 2, 1, x; m, n) + (0, 3, 2, x, 1, z; m, n)\}, \tag{16}$$

where $z = 1$.

The $\lim_{n\to\infty} c(3, m, n)(0, 3, 2, 1, x/n; m, n)$ and $\lim_{n\to\infty} c(3, m, n)(0, 3, 2, x/n, 1, z; m, n)$ are given by (12) and (15) respectively. Substituting these results in (16) and simplifying we get

$$\begin{aligned}\lim_{n\to\infty} \Pr(n\theta_2 \le x) = \frac{2^{2m+3}}{\Gamma(m+1)\Gamma(2m+3)} \Big\{ 2 \int_0^\infty u^m e^{-u}\, du \\ \cdot \int_0^x u^{2m+3} e^{-2u}\, du - 2\int_0^\infty u^{m+1} e^{-u}\, du \int_0^x u^{2m+2} e^{-2u}\, du \\ - 4x^{m+2}e^{-x}\int_0^x u^{2m+1}e^{-2u}\, du + 2x^{2m+3}e^{-2x}\int_0^x u^m e^{-u}\, du \\ + x^{m+2}e^{-x}\left[\int_x^\infty u^{m+1}e^{-u}\, du \int_0^x u^m e^{-u}\, du - \int_x^\infty u^m e^{-u}\, du \int_0^x u^{m+1} e^{-u}\, du\right]\Big\},\end{aligned}$$

Or,

$$\begin{aligned}\lim_{n\to\infty} \Pr(n\theta_2 \le x) = \frac{2^{2m+3}}{\Gamma(m+1)\Gamma(2m+3)} \Big\{ 2\Gamma(m+1) \int_0^x u^{2m+3} e^{-2u}\, du \\ - 2\Gamma(m+2)\int_0^x u^{2m+2}e^{-2u}\, du - 4x^{m+2}e^{-x}\int_0^x u^{2m+1}e^{-2u}\, du + 2x^{2m+3}e^{-2x} \\ \cdot \int_0^x u^m e^{-u}\, du + x^{m+2}e^{-x}\left[\int_x^\infty u^{m+1}e^{-u}\, du \int_0^x u^m e^{-u}\, du \right.\\ \left. - \int_x^\infty u^m e^{-u}\, du \int_0^x u^{m+1}e^{-u}\, du\right]\Big\}.\end{aligned}$$

Thus the limiting distribution of any intermediate root can be obtained by the above method.

7. Further problems. The limiting distribution of the largest root is found to be very helpful in obtaining the distribution of the sum of roots when $m = 0$. This condition implies that when the results are applied to canonical correlations the numbers of variates in the two sets differ by unity. The distributions for the sum of roots have been derived under the above condition for $l = 2, 3$ and 4 and the results are being presented in the next paper of this series.

8. Acknowledgements. I am extremely thankful to Drs. P. L. Hsu and Harold Hotelling for guidance and help in this research.

REFERENCES

[1] P. L. Hsu, "On the distribution of roots of certain determinantal equations," *Annals of Eugenics*, Vol. 3 (1939), pp. 250–258.

[2] D. N. Nanda, "Distribution of a root of a determinantal equation," *Annals of Math. Stat.*, Vol. 18 (1947).

[3] S. N. Roy, "The individual sampling distribution of the maximum, the minimum and any intermediate of the 'p' statistics on the null hypothesis," *Sankhyā*, Vol. 8 (1943).

ON A SOURCE OF DOWNWARD BIAS IN THE ANALYSIS OF VARIANCE AND COVARIANCE

By William G. Madow

Institute of Statistics, University of North Carolina

1. Summary. It is shown that if, in the analysis of variance, the experiments are not in a state of statistical control due to variations in the true means, then the test will have a downward bias. The power function of the analysis of variance test is obtained when this downward bias is present.

2. Introduction. To introduce the discussion of this bias let us consider the generalized Student's hypothesis.

Let $y_1, \cdots, y_{kN}$ be normally and independently distributed with variance σ^2, and let the expected value of $y_{i\nu}$, be $a_{i\nu}$.[1] Then the generalized Student's hypothesis is

(Null hypothesis) $$a_{i\nu} = a$$

and the class of alternative hypotheses against which the null hypothesis is tested is

(Class A) $$a_{i\nu} = a_i\,.$$

From the statement of the null hypothesis and the alternatives of Class A it follows that both the null hypothesis and the alternatives of Class A require that

$$a_{i1} = \cdots = a_{iN}\,. \tag{1.1}$$

Since our experiments are rarely in such perfect statistical control that (1.1) holds whether or not the null hypothesis is true, it becomes reasonable to investigate the existing F test when instead of the alternatives to the null hypothesis being of Class A, they are simply Class B:

(Class B) Equation (1.1) is false for at least one value of i.

Furthermore, for many practical purposes we would prefer to test the average null hypothesis:

(Average null hypothesis) $$\bar{a}_i = \bar{a},$$

where $N\bar{a}_i = a_{i1} + \cdots + a_{iN}$ and $k\bar{a} = \bar{a}_1 + \cdots + \bar{a}_k$, instead of the null hypothesis, the alternatives to the average null hypothesis being of Class C.

(Class C) The $a_{i\nu}$ can have any values such that not all the $\bar{a}_i$ equal $\bar{a}$.

[1] Throughout this paper the letter i will assume all integral values from 1 to k, the letters μ, ν will assume all integral values from 1 to N, the letters γ, η will assume all integral values from 1 to m, the letter α will assume all integral values from $n_1 + \cdots + n_{\gamma-1} + 1$ to $n_1 + \cdots + n_\gamma$, $(n_0 = 0)$, and α_1, α_2 will assume all integral values from 0 to ∞.

The F-test of the null hypothesis against the alternatives of Class A is, as is well known,

$$F = \frac{k(N-1)\sum_i (\bar{y}_i - \bar{y})^2}{(k - 1 \sum_{i,\nu} (y_{i\nu} - \bar{y}_i)^2)}$$

where $N\bar{y}_i = y_{i1} + \cdots + y_{iN}$ and $k\bar{y} = \bar{y}_1 + \cdots + \bar{y}_k$. To answer the questions formulated above concerning the F-test when the average null hypothesis or the alternatives of classes B or C are true, we must then calculate the distribution of F under these various conditions. This is done in Section 3.

A somewhat informal means of obtaining the conclusions is that of studying F itself. Taking the expected values of the numerator and denominator of F and defining

$$\phi_1^2 = \frac{N\sum_i (\bar{a}_i - \bar{a})^2}{(k-1)\sigma^2}$$

$$\phi_2^2 = \frac{1}{k(N-1)\sigma^2}\sum_{i,\nu} (a_{i\nu} - \bar{a}_i)^2$$

we obtain as the ratio of the two expected values

$$\bar{F} = \frac{1+\phi_1^2}{1+\phi_2^2}.$$

It is well known that, in general, the larger the value of N the more closely will F approximate $\bar{F}$. From this fact it is easy to see why if the null hypothesis is true, then $F \sim 1$, whereas if the null hypothesis is false but an alternative of Class A is true then

$$F \sim 1 + \phi_1^2 > 1$$

so that large values of F become more likely than if the null hypothesis were true. However, if an alternative of Class B is true then

$$F \sim \frac{1+\phi_1^2}{1+\phi_2^2}$$

so that if $\phi_1^2 < \phi_2^2$, smaller values of F occur more frequently than indicated by the null hypothesis. Thus we would tend to accept the null hypothesis more frequently than desired when it is false. Even when the null hypothesis is false so that $\phi_1^2 > 0$, the values of F will tend to be less if $\phi_2^2 > 0$ than if $\phi_2^2 = 0$ whether or not $\phi_1^2 < \phi_2^2$. Not only is the probability of an error of the first kind less than the value ϵ we may have previously selected, but also the power of the test is less than would be indicated by Tang's tables [1]. The lack of statistical control represented by variation of expected values within a class has the effect of making it less likely than the standard F-test indicates that the null

hypothesis will be rejected whether it be true or false. Furthermore, even for relatively low values of ϕ_2^2, the reductions in the probabilities of rejection may be over 40 per cent as indicated by some examples given below.

If the average null hypothesis is true but (1.1) is false it follows that

$$F \sim \frac{1}{1 + \phi_2^2},$$

so that the full effect of the downward bias occurs in that case. Thus in cases where statistical control is lacking, to test the average null hypothesis by the F-test may well result in accepting the hypothesis when it is false. If the null hypothesis is rejected, however, then we can expect that the differences among the true means are even larger than indicated by Tang's tables.

To illustrate, it is shown in Section 4 that if $k = 5$ and $N = 7$, then the probability of rejecting the average null hypothesis when it is true, but (1.1) is false will not be the preassigned .05 but something less than .03 if $\phi_2^2 > .05$. Furthermore, if $\phi_2^2 > .07$, then the power of the F tests for this example will be reduced by at least 40 per cent whatever the value of ϕ_1^2.

The conclusions reached above remain valid for the analysis of variance and covariance in general. In the general case however, the value of the average null hypothesis in simplifying the analysis may be considerably reduced since the parameter ϕ_1^2 no longer vanishes when the average null hypothesis is true. For example, if $Ey_\nu = \beta_\nu x_\nu$, and if the average null hypothesis is $\bar{\beta} = 0$, where $N\bar{\beta} = \beta_1 + \cdots + \beta_N$, then upon calculating

$$\phi_1^2 = \frac{\left(\sum_\nu \beta_\nu x_\nu^2\right)^2}{\sigma^2 \sum_\nu x_\nu^2}$$

we see that ϕ_1^2 will not vanish in general if $\bar{\beta}$ vanishes.

Although as shown above the average null hypothesis may not have too great importance in the case of regression, yet if the "variance between treatments" is a function of arithmetic means of the random variables as in the "pure" analysis of variance the average null hypothesis may well be very useful. Simple examples of this are provided by the randomized block, Latin square, and similar designs.

The distributions that we shall need are given in Section 3. The inequalities on the basis of which the bias is demonstrated are obtained in Section 4.

It would be highly desirable to have Tang's tables extended so that they might provide the answers to the questions raised by this source of bias. In the absence of such extensions the inequalities of Section 4 may give some rough idea, but these inequalities are not sharp enough.

3. The calculation of the distributions. The following theorem was proved, although not explicitly stated, as part of an earlier note [2]. (Note the change from x_i to y_i as the notation for the random variable.)

THEOREM 1. *Let $y_1, \cdots, y_N$ be normally and independently distributed with variance σ^2 and means $a_1, \cdots, a_N$ and let $q_1, \cdots, q_m$ be quadratic forms*

$$q_\gamma = \sum_{\mu,\nu} a_{\mu\nu}^{(\gamma)} y_\mu y_\nu$$

in $y_1, \cdots, y_N$ of ranks $n_1, \cdots, n_m$. Then, if an orthogonal transformation

$$y_\nu = \sum_\mu c_{\nu\mu} z_\mu$$

exists such that

$$q_\gamma = \sum_\alpha z_\alpha^2, \tag{2.1}$$

it follows that the random variables q_γ/σ^2 are independently distributed in χ'^2 distributions with degrees of freedom $n_1, \cdots, n_m$ and parameters $\lambda_1, \cdots, \lambda_m$, where

$$\lambda_\gamma = \frac{1}{2\sigma^2} \sum_{\mu,\nu} a_{\mu\nu}^{(\gamma)} a_\mu a_\nu = \frac{Eq_\gamma}{2\sigma^2} - \frac{n_\gamma}{2}.$$

Various conditions for the existence of an orthogonal transformation satisfying (2.1) of Theorem 1 have been given. Among these are:

1. *Cochran's* [3] *condition.* If $\sum_\gamma q_\gamma = \sum_\nu y_\nu^2$ then a necessary and sufficient condition for the existence of an orthogonal transformation satisfying (2.1) is $\sum_\gamma n_\gamma = N$.

2. *Craig's* [4] *condition.* If A_γ denotes the matrix $(a_{\mu\nu}^{(\gamma)})$ then a necessary and sufficient condition for the existence of an orthogonal transformation satisfying (2.1) is $A_\gamma A_\eta = \delta_{\gamma\eta} A_\gamma$ where $\delta_{\gamma\eta}$ is the null matrix if $\gamma \neq \eta$ and the identity matrix if $\gamma = \eta$.

3. *Linear Hypothesis condition.* (Kolodziejczyk [5]) If λ be the likelihood ratio test of a linear hypothesis and if $E^2 = 1 - \lambda^{2/N}$, then $E^2 = q_1/(q_1 + q_2)$ and an orthogonal transformation exists satisfying (2.1) with $m = 2$.

To summarize some results obtained by Tang [1], let us state

THEOREM 2. *If $\chi_1'^2$ and $\chi_2'^2$ are independently distributed in distributions with n_1 and n_2 degrees of freedom and parameters λ_1 and λ_2, and if*

$$E^2 = \frac{\chi_1'^2}{\chi_1'^2 + \chi_2'^2},$$

then the probability density of E^2 is

$$\begin{aligned} p = p(E^2 \mid \lambda_1, \lambda_2, n_1, n_2) &= e^{-\lambda_1-\lambda_2}(E^2)^{(n_1/2)-1}(1 - E^2)^{(n_2/2)-1} \\ &\cdot \sum_{\alpha_1,\alpha_2} \frac{\lambda_1^{\alpha_1}\lambda_2^{\alpha_2}\Gamma\left(\frac{n_1+n_2}{2}+\alpha_1+\alpha_2\right)}{\alpha_1!\,\alpha_2!\,\Gamma\left(\frac{n_1}{2}+\alpha_1\right)\Gamma\left(\frac{n_2}{2}+\alpha_2\right)}(E^2)^{\alpha_1}(1-E^2)^{\alpha_2}. \end{aligned} \tag{2.2}$$

By assigning certain values to λ_1 and λ_2 we obtain the following special cases of (2.2)

$$p_1 = p(E^2 \mid \lambda_1, 0, n_1, n_2) = e^{-\lambda_1}(E^2)^{(n_1/2)-1}(1 - E^2)^{(n_2/2)-1} \cdot \sum_{\alpha_1} \frac{\lambda_1^{\alpha_1}\Gamma\left(\frac{n_1+n_2}{2}+\alpha_1\right)}{\alpha_1!\,\Gamma\left(\frac{n_1}{2}+\alpha_1\right)\Gamma\left(\frac{n_2}{2}\right)}(E^2)^{\alpha_1} \tag{2.3}$$

$$p_2 = p(E^2 \mid 0, \lambda_2, n_1, n_2) = e^{-\lambda_2}(E^2)^{(n_2/2)-1}(1 - E^2)^{(n_2/2)-1} \cdot \sum_{\alpha_2} \frac{\lambda_2^{\alpha_2}\Gamma\left(\frac{n_1+n_2}{2}+\alpha_2\right)}{\alpha_2!\,\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}+\alpha_2\right)}(1 - E^2)^{\alpha_2} \tag{2.4}$$

$$p_0 = p(E^2 \mid 0, 0, n_1, n_2) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}(E^2)^{(n_1/2)-1}(1 - E^2)^{(n_2/2)-1}. \tag{2.5}$$

It is noted that (2.3) is Tang's distribution (112) upon which the calculations of his tables were based. To see this we need only make the correspondence

This paper	*Tang*
λ_1	λ
n_1, n_2	f_1, f_2
α_1	i

We define ϵ to be the probability of an error of the first kind. Tang obtained the critical values E_ϵ^2 of E^2 by requiring that

$$P_I = \int_{E_\epsilon^2}^1 p_0 \, dE^2$$

$$= \epsilon \qquad \text{say .01 or .05.}$$

Then he calculated

$$P_{II} = \int_0^{E_\epsilon^2} p_1(E^2 \mid \lambda_1, 0, n_1, n_2) \, dE^2$$

using the values of E_ϵ^2 obtained above. Hence $1 - P_{II}$ is the power of the test.

If, however, $\lambda_1 = 0$ but $\lambda_2 \neq 0$, then to find

$$P_{III} = \int_{E_\epsilon^2}^1 p_2(E^2 \mid 0, \lambda_2, n_1, n_2) \, dE^2$$

we could make the transformation $G^2 = 1 - E^2$ and find

$$P_{III} = \int_0^{1-E_\epsilon^2} p(G^2 \mid 0, \lambda_2, n_1, n_2)\, dG^2.$$

It is easy to verify that

$$p(G^2 \mid 0, \lambda_2, n_1, n_2) = p_1(E^2 \mid \lambda_2, 0, n_2, n_1)$$

if we put G in place of E^2 in the latter density. It follows that to calculate P_{III} it would be sufficient to have full tables of Tang's distribution since

$$P_{III} = \int_0^{1-E_\epsilon^2} p_1(E^2 \mid \lambda_2, 0, n_2, n_1)\, dE^2.$$

Tang's tables are not however sufficiently extensive. Furthermore, tables of (2.2) are also necessary. As yet these tables do not exist. However, some useful conclusions can be drawn from the inequalities obtained in the following section.

First, however, let us evaluate n_1, n_2, λ_1 and λ_2 for the generalized Student's hypothesis discussed in the introduction. It is easy to see that $n_1 = k - 1$ and $n_2 = k(N - 1)$. To evaluate λ_1 and λ_2 we note from Theorem 1 that we only need substitute Ey_{ij} for y_{ij} in q_1 and q_2 where

$$q_1 = N \sum_i (\bar{y}_i - \bar{y})^2$$

$$q_2 = \sum_{i,\nu} (y_{i\nu} - \bar{y}_i)^2.$$

Upon making these substitutions we obtain

$$\lambda_1 = \frac{N}{2\sigma^2} \sum_i (\bar{a}_i - \bar{a})^2$$

$$\lambda_2 = \frac{1}{2\sigma^2} \sum_{i,\nu} (a_{i\nu} - \bar{a}_i)^2.$$

Thus the various hypotheses concerning the a_{ij} influence the distribution of F or $E^2 = 1/(1 + Fn_1/n_2)$ by affecting the values of λ_1 and λ_2.

4. Limits of the values of p. It follows readily from (2.2) that,

$$p = \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} (E^2)^{(n_1/2)-1}(1 - E^2)^{(n_2/2)-1} \tag{3.1}$$

$$\cdot e^{-\lambda_1-\lambda_2} \sum_{\alpha_1 \cdot \alpha_2} \frac{\lambda_1^{\alpha_1}\lambda_2^{\alpha_2}}{\alpha_1!\,\alpha_2!} (E^2)^{\alpha_1}(1 - E^2)^{\alpha_2} C_{\alpha_1\alpha_2}$$

where

$$C_{\alpha_1\alpha_2} = \frac{\Gamma\left(\frac{n_1 + n_2}{2} + \alpha_1 + \alpha_2\right)\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2} + \alpha_1\right)\Gamma\left(\frac{n_2}{2} + \alpha_2\right)\Gamma\left(\frac{n_1 + n_2}{2}\right)}.$$

Now if $a > 0, b > 0$, and j is an integer > 1, we have

$$\left(1 + \frac{a}{b}\right)\left(1 + \frac{a}{b+2}\right) \cdots \left(1 + \frac{a}{b + 2(j-1)}\right) < \left(1 + \frac{a}{b}\right)^j.$$

Hence, it follows that

$$1 \leq C_{\alpha_1\alpha_2} \leq \left(\frac{n_1 + n_2}{n_1}\right)^{\alpha_1} \left(\frac{n_1 + n_2 + 2\alpha_1}{n_2}\right)^{\alpha_2}.$$

Substituting we see that

$$p_0 e^{-\lambda_1-\lambda_2} \cdot e^{\lambda_1 E^2 + \lambda_2(1-E^2)} \leq p \leq p_0 e^{-\lambda_1-\lambda_2} \tag{3.2}$$
$$\cdot \exp\left\{\lambda_1 E^2 \left(\frac{n_1 + n_2}{n_1}\right) \exp\left[\frac{2\lambda_2(1 - E^2)}{n_2}\right] + \lambda_2(1 - E^2)\left(\frac{n_1 + n_2}{n_2}\right)\right\}$$

and

$$p_1 e^{-\lambda_2 + \lambda_2(1-E^2)} < p < p_1 \exp\left[-\lambda_2 + \lambda_2(1 - E^2)\left(\frac{n_1 + n_2}{n_2}\right) + 2\frac{\lambda_2}{n_2}\right]. \tag{3.3}$$

Let $2n_i\phi_i^2 = \lambda_i$, $i = 1, 2$.

THEOREM 3. *Let* $\epsilon = \int_{E_\epsilon^2}^1 p_0 \, dE^2$ *so that* ϵ *is the probability of an error of the first kind. Then, for all values of* ϕ_2^2

$$\epsilon > \int_{E_\epsilon^2}^1 p_2 \, dE^2 \tag{3.4}$$

and if $E^2 > n_1/(n_1 + n_2)$, *it follows that*

$$\epsilon > \epsilon \exp\{-2n_2\phi_2^2 + 2\phi_2^2(1 - E_\epsilon^2)(n_1 + n_2)\} > \int_{E_1^2}^1 p_2 \, dE^2 > \epsilon e^{-n_2\phi_2^2}. \tag{3.5}$$

Furthermore, for all values of ϕ_2^2

$$\int_{E_\epsilon^2}^1 p_1 \, dE^2 > \int_{E_\epsilon^2}^1 p \, dE^2, \tag{3.6}$$

and if $E^2 > (n_1 + 2)/(n_1 + n_2)$, *it follows that*

$$\int_{E_\epsilon^2}^1 p_1 \, dE_2 > \exp\{-2n_2\phi_2^2 + 2\phi_2^2(1 - E_\epsilon^2)(n_1 + n_2) \, 2\phi_2^2\} \int_{E_\epsilon^2}^1 p_1 \, dE^2 \tag{3.7}$$
$$> \int_{E_\epsilon^2}^1 p \, dE^2 > e^{-2n_2\phi_2^2} \int_{E_\epsilon^2}^1 p_1 \, dE^2.$$

Finally, if γ *can assume the two values* 0 *and* 2, *it follows that if*

$$\phi_2^2 > \frac{-\log \delta}{2(E_\epsilon^2(n_1 + n_2) - (n_1 + \gamma))} > 0, \tag{3.8}$$

then if $\gamma = 0$,

$$\int_{E_\epsilon^2}^1 p_2 \, dE^2 < \epsilon\delta \tag{3.9}$$

and if $\gamma = 2$

$$\int_{E_\epsilon^2}^1 p\, dE^2 < \delta \int_{E_\epsilon^2}^1 p_1\, dE^2. \tag{3.10}$$

PROOF. To prove (3.4) and (3.6) it is only necessary to follow Daly's [6] procedure.[2] Since

$$\exp\{-2n_2\phi_2^2 + 2\phi_2^2(1 - E^2)(n_1 + n_2) + \gamma\phi_2^2\}$$

and

$$\exp\{-n_2\phi_2^2E^2\}$$

are decreasing functions of E^2, and

$$\exp\{-2n_2\phi_2^2 + 2\phi_2^2(1 - E^2)(n_1 + n_2) + \gamma\phi_2^2\} < 1$$

if

$$E^2 > \frac{n_1 + \gamma}{n_1 + n_2}$$

the inequalities (3.5) and (3.7) follow immediately from (3.2) and (3.3). Finally

$$\exp\{-2n_2\phi_2^2 + 2\phi_2^2(1 - E^2)(n_1 + n_2) + \gamma\phi_2^2\} < \delta < 1$$

if (3.8) is true, so that (3.9) and (3.10) follow.

From (3.8), (3.9) and (3.10) we can calculate either a lower limit for the bias, if we know ϕ_2, or the upper limit that ϕ_2 can have if we wish the bias to be not greater than some given amount. Thus these limits do not answer the important question of what is a value ϕ_2 such that if $\phi_2 < \phi$ then the bias is less than $(1 - \delta)\epsilon$. They only provide a value ϕ' of ϕ_2 such that if $\phi_2 > \phi'$ then the bias is at least $(1 - \delta)\epsilon$.

If, for example, $\delta = .5$ and $n_1 = 1$ as in the case of Students' ratio; we have if $\gamma = 0$

$$\phi_2^2 > \frac{.693}{2(n_2E^2 - 1)}$$

and if $\epsilon = .05$, then E^2 decreases steadily from .903 if $n_2 = 2$, to .063 if $n_2 = 60$' and the corresponding lower limits of ϕ_2^2 decrease from .43 to .12. Thus, if $\phi_2^2 > .43$ or .12 in these two cases, it follows that the probability of rejecting the average null hypothesis will be not .05 but something less than .025.

If $\delta = .6$ and $n_1 = 4$, $n_2 = 30$ then we can evaluate the lower limit of ϕ_2^2 for the example given in the introduction finding.

$$\phi_2^2 > \frac{.511}{2(.279)(34) - 8} = .05$$

implies a downward bias of at least 40 per cent of .05. Also, if $\phi_2^2 > .07$ then for

[2] The procedure followed is given in [6] on pp. 4, 5, equations (2.2) through Lemma 1.

any value of ϕ_1 the power of the analysis of variance test is reduced at least 40 per cent.

5. Conclusions. The rather sharp effects of a moderate lack of statistical control on the probabilities associated with the F-test indicates the importance of testing for statistical control outside of the industrial applications now made. Furthermore, it would seem advisable to investigate tests and designs that are less sensitive to the lack of control than is the F-test.

REFERENCES

[1] P. C. Tang, "The power function of the analysis of variance tests with tables and illustrations of their use," *Statistical Research Memoirs*, Vol. 2 (1938), pp. 126–157.

[2] William G. Madow, "The distribution of quadratic forms in non-central normal random variables," *Annals of Mathem. Stat.*, Vol. 9 (1940), pp. 100–104.

[3] W. G. Cochran, "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance," *Cambridge Phil. Soc. Proc.* Vol. 30 (1934), pp. 178–191.

[4] A. T. Craig, "Note on the independence of certain quadratic forms," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 195–197.

[5] S. Kolodziejczyk, "On the important class of statistical hypotheses," *Biometrika*, Vol. 27 (1935), pp. 161–190.

[6] J. F. Daly, "On the unbiased character of likelihood-ratio tests for independence in normal systems," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 1–33. The procedure followed is given on pp. 4, 5, equations (2.2) through Lemma 1.

MIXTURE OF DISTRIBUTIONS

By Herbert Robbins

Department of Mathematical Statistics, University of North Carolina

1. Summary. Mixtures of measures or distributions occur frequently in the theory and applications of probability and statistics. In the simplest case it may, for example, be reasonable to assume that one is dealing with the mixture in given proportions of a finite number of normal populations with different means or variances. The mixture parameter may also be denumerably infinite, as in the theory of sums of a random number of random variables, or continuous, as in the compound Poisson distribution.

The operation of Lebesgue-Stieltjes integration, $\int f(x)\, d\mu$, is linear with respect to both integrand $f(x)$ and measure μ. The first type of linearity has as its continuous analog the theorem of Fubini on interchange of order of integration; the second type of linearity has a corresponding continuous analog which is of importance whenever one deals with mixtures of measures or distributions, and which forms the subject of the present paper. Other treatments of the same subject have been given ([1], [2]; see also [3], [4]) but it is hoped that the discussion given here will be useful to the mathematical statistician.

A general measure theoretic form of the fundamental theorem is given in Section 2, and in Section 3 the theorem is formulated in terms of finite dimensional spaces and distribution functions. The operation of convolution as an example of mixture is treated briefly in Section 4, while Section 5 is devoted to random sampling from a mixed population.

We shall refer to *Theory of the Integral* by S. Saks (second edition, Warszawa, 1937) as [*S*], and the *Mathematical Methods of Statistics* by H. Cramér (Princeton, 1946) as [*C*].

2. Mixture of measures in general. Let $X(Y)$ be a space with points $x(y)$ and let $\mathfrak{X}(\mathfrak{Y})$ be a σ-field of subsets of $X(Y)$. Let ν be a measure on $\mathfrak{Y}$. Let μ_y be for a. e. (ν) y a measure on $\mathfrak{X}$, such that $\mu_y(S)$ is for every S in $\mathfrak{X}$ a measurable $(\mathfrak{Y})$ function of y. Define for every S in $\mathfrak{X}$,

$$\mu(S) = \int_Y \mu_y(S)\, d\nu. \tag{1}$$

Theorem 1. *μ is a measure on $\mathfrak{X}$. If $\nu(Y) = \mu_y(X) = 1$, then $\mu(X) = 1$.*

Proof. Clear.

Theorem 2. *If $f(x)$ is any non-negative or non-positive function measurable $(\mathfrak{X})$ then the function*

$$g(y) = \int_X f(x)\, d\mu_y \tag{2}$$

is measurable ($\mathfrak{Y}$), and

$$\int_X f(x)\, d\mu = \int_Y g(y)\, d\nu. \tag{3}$$

Proof. First let $f_0(x)$ be any non-negative simple function [S, p. 7] of the form

$$f_0(x) = \{a_1, S_1; \cdots ; a_k, S_k\} \tag{4}$$

where the S_i are disjoint sets in $\mathfrak{X}$ such that $X = \sum_1^k S_i$ and the a_i are non-negative constants. Then

$$g_0(y) = \int_X f_0(x)\, d\mu_y = \sum_1^k a_i \mu_y(S_i) \tag{5}$$

is a non-negative function measurable ($\mathfrak{Y}$), and from (1) it follows that each side of (3) is equal to $\sum_1^k a_i\mu(S_i)$. Hence the theorem holds in this case.

Next let $f(x)$ be any non-negative function measurable ($\mathfrak{X}$); then [S, p. 14] there exists a sequence $f_n(x)$ of simple functions such that for every x,

$$0 \leq f_1(x) \leq f_2(x) \leq \cdots ; \qquad \lim_{n\to\infty} f_n(x) = f(x). \tag{6}$$

Setting

$$g_n(y) = \int_X f_n(x)\, d\mu_y, \qquad g(y) = \int_X f(x)\, d\mu_y, \tag{7}$$

it follows from the theorem of monotone convergence [S, p. 28] and from the preceding paragraph that

$$\int_X f(x)\, d\mu = \lim_{n\to\infty} \int_X f_n(x)\, d\mu = \lim_{n\to\infty} \int_Y g_n(y)\, d\nu, \tag{8}$$

$$g(y) = \lim_{n\to\infty} \int_X f_n(x)\, d\mu_y = \lim_{n\to\infty} g_n(y). \tag{9}$$

From (6) and (9) it follows that for a.e. $(\nu)y$,

$$0 \leq g_1(y) \leq g_2(y) \leq \cdots ; \qquad \lim_{n\to\infty} g_n(y) = g(y). \tag{10}$$

Hence $g(y)$ is measurable ($\mathfrak{Y}$), and from the theorem of monotone convergence,

$$\int_Y g(y)\, d\nu = \lim_{n\to\infty} \int_Y g_n(y)\, d\nu. \tag{11}$$

Equation (3) now follows from (8) and (11).

By passing from $f(x)$ to $-f(x)$ we establish (3) when $f(x)$ is any non-positive function measurable ($\mathfrak{X}$). This completes the proof of Theorem 2.

If $f(x)$ is an arbitrary function measurable ($\mathfrak{X}$) we define

$$f^+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0 \\ 0 & \text{otherwise} \end{cases}, \qquad f^-(x) = \begin{cases} f(x) & \text{if } f(x) \leq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{12}$$

so that

$$f(x) = f^+(x) + f^-(x) \tag{13}$$

is the sum of two functions measurable ($\mathfrak{X}$) of constant sign. By Theorem 2 the functions

$$g_1(y) = \int_X f^+(x)\, d\mu_y, \qquad g_2(y) = \int_X f^-(x)\, d\mu_y \tag{14}$$

are measurable ($\mathfrak{Y}$) and

$$0 \leq \int_X f^+(x)\, d\mu = \int_Y g_1(y)\, d\nu \leq \infty, \tag{15}$$

$$0 \geq \int_X f^-(x)\, d\mu = \int_Y g_2(y)\, d\nu \geq -\infty. \tag{16}$$

The integral $\int_X f(x)\, d\mu$ exists if and only if at least one of the two quantities (15) and (16) is finite [S, p. 20].

THEOREM 3. *A necessary and sufficient condition that*

$$\int_X f(x)\, d\mu = \int_Y \left\{ \int_X f(x)\, d\mu_y \right\} d\nu \tag{17}$$

is that at least one of the two quantities (15) *and* (16) *be finite.*

PROOF. By the remark preceding Theorem 3 the condition is clearly necessary. Now suppose, e.g., that (15) is finite; we must show that (17) holds. By hypothesis,

$$\int_X f^+(x)\, d\mu < \infty, \qquad \int_X f(x)\, d\mu = \int_X f^+(x)\, d\mu + \int_X f^-(x)\, d\mu. \tag{18}$$

From (18) and (15) it follows that $0 \leq g_1(y) < \infty$ for a.e. $(\nu)y$; hence

$$\int_X f(x)\, d\mu_y = \int_X f^+(x)\, d\mu_y + \int_X f^-(x)\, d\mu_y = g_1(y) + g_2(y) \tag{19}$$

exists for a.e. $(\nu)y$. From the finiteness of (15) it follows that

$$\int_Y (g_1(y) + g_2(y))\, d\nu = \int_Y g_1(y)\, d\nu + \int_Y g_2(y)\, d\nu \tag{20}$$

exists. Hence from (19), the integral

$$\int_Y \left\{ \int_X f(x)\, d\mu_y \right\} d\nu = \int_Y (g_1(y) + g_2(y))\, d\nu \tag{21}$$

exists. Equation (17) now follows from (21), (20), (15), and (18). This completes the proof of Theorem 3.

COROLLARY 1. *If $\mu(X) < \infty$, and if $f(x)$ is bounded from above or from below, then both sides of* (17) *exist and the equality holds.*

PROOF. If, say, $f(x) \leq C < \infty$, then

$$0 \leq \int_X f^+(x)\, d\mu \leq C \cdot \mu(X) < \infty,$$

and the result follows from Theorem 3.

We shall now show by an example that the existence and even the finiteness of the right side of (17) does not imply the existence of the left side.

Let $X = Y = \{1, 2, \cdots, n, \cdots\}$ and let $\mathfrak{X}(\mathfrak{Y})$ consist of all subsets of $X(Y)$. Let ν be the measure which assigns mass c_n to n, where the c_n are positive constants such that $\sum_1^\infty c_n = 1$. Let μ_n assign the mass $1/2n$ to each of the points $1, 2, \cdots, 2n$. Let $f(x)$ be such that $f(1) = b_1, f(2) = -b_1, f(3) = b_2, f(4) = -b_2, \cdots$ where the b_n are positive constants. Then

$$\int_X f(x)\, d\mu_n = 0 \qquad (n = 1, 2, \cdots),$$

so that

$$\int_Y \left\{ \int_X f(x)\, d\mu_n \right\} d\nu = 0.$$

The measure μ defined by (1) assigns to each n a positive value $\mu(n)$ given by

$$\mu(1) = \mu(2) = c_1 \cdot (2)^{-1} + c_2 \cdot (2 \cdot 2)^{-1} + c_3 \cdot (2 \cdot 3)^{-1} + \cdots$$

$$\mu(3) = \mu(4) = c_2 \cdot (2 \cdot 2)^{-1} + c_3 \cdot (2 \cdot 3)^{-1} + \cdots$$

$$\cdots$$

where $\mu(X) = \sum_1^\infty \mu(n) = \sum_1^\infty c_n = 1$.

Now fix the b_n and c_n in such a way that

$$b_1 \cdot \mu(1) + b_2 \cdot \mu(3) + b_3 \cdot \mu(5) + \cdots = \infty.$$

Then

$$\int_X f^+(x)\, d\mu = -\int_X f^-(x)\, d\mu = \infty,$$

so that the left side of (17) does not exist, even though $\nu(Y) = \mu_y(X) = \mu(X) = 1$ and the right side of (17) exists and is equal to zero.

3. A restatement of the preceding results in the form most useful in probability theory. Let $x = (x_1, \cdots, x_n)$ be a point in the n-dimensional Euclidean space R_n, and let B_n denote the σ-field of Borel sets in R_n. Let S_x denote the half-open interval in R_n consisting of all points $(w_1, \cdots, w_n)$ in R_n satisfying the inequalities

$$w_1 \leq x_1, \cdots, w_n \leq x_n; \tag{22}$$

then if μ is any probability measure on B_n the function

$$F(x) = \mu(S_x) \tag{23}$$

is the distribution function corresponding to μ. Conversely, if $F(x)$ is any distribution function in R_n [C, p. 80] there is a unique probability measure μ on B_n such that (23) holds. As a matter of notation we write for any Borel measurable $f(x)$,

$$\int_{R_n} f(x)\, d\mu = \int_{-\infty}^{\infty} f(x)\, dF(x) \tag{24}$$

provided the integral on the left exists.

Now let $y = (y_1, \cdots, y_m)$ be a point in R_m, let $G(y)$ be a distribution function, and let ν denote the corresponding probability measure on B_m. Let $F(x,y)$ be for a.e. $(\nu)y$ a distribution function in x, and for every x a Borel measurable function of y, and let μ_y be the corresponding probability measure on B_n.

THEOREM 4. *The function*

$$H(x) = \int_{-\infty}^{\infty} F(x, y)\, dG(y) \tag{25}$$

is a distribution function in R_n. Let μ denote the corresponding probability measure on B_n. Then for any S in B_n, $\mu_y(S)$ is a Borel measurable function of y and

$$\mu(S) = \int_{-\infty}^{\infty} \mu_y(S)\, dG(y). \tag{26}$$

PROOF. Let C denote the class of all Borel sets S in R_n such that $\mu_y(S)$ is a Borel measurable function of y. We shall show that C is a normal class [S, p. 83].

(i) If $S_1, S_2, \cdots$ is a sequence of disjoint sets in C and if $S = \sum_1^{\infty} S_n$, then

$$\mu_y(S) = \mu_y\left(\sum_1^{\infty} S_n\right) = \sum_1^{\infty} \mu_y(S_n)$$

is a convergent series of Borel measurable functions and is therefore itself a Borel measurable function.

(ii) If $S_1 \supset S_2 \supset \cdots$ is a decreasing sequence of sets in C and if $S = \prod_1^{\infty} S_n$, then

$$\mu_y(S) = \mu_y\left(\prod_1^{\infty} S_n\right) = \lim_{n \to \infty} \mu_y(S_n)$$

is the limit of a sequence of Borel measurable functions and is therefore a Borel measurable function.

Hence C is a normal class. But C contains every interval S_x, for $\mu_y(S_x) = F(x, y)$ was assumed to be a Borel measurable function of y for every x. It follows [S, p. 85] that $C = B_n$.

It now follows from Theorem 1 that the set function $\mu(S)$ defined by (26) is a probability measure on B_n. The corresponding distribution function is the function $H(x)$ defined by (25). Thus Theorem 4 is proved.

Let $f(x) = f^+(x) + f^-(x)$ be any Borel measurable function. Then from Theorem 2, the integrals

$$\int_{-\infty}^{\infty} f^+(x)\, dH(x) = \int_{-\infty}^{\infty} f^+(x)\, d_x \left\{ \int_{-\infty}^{\infty} F(x, y)\, dG(y) \right\} \tag{27}$$
$$= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f^+(x)\, d_x F(x, y) \right\} dG(y),$$

$$\int_{-\infty}^{\infty} f^-(x)\, dH(x) = \int_{-\infty}^{\infty} f^-(x)\, d_x \left\{ \int_{-\infty}^{\infty} F(x, y)\, dG(y) \right\} \tag{28}$$
$$= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f^-(x)\, d_x F(x, y) \right\} dG(y)$$

exist. The following theorem is an immediate consequence of Theorem 3 and Corollary 1.

THEOREM 5. *A necessary and sufficient condition that*

$$\int_{-\infty}^{\infty} f(x)\, d_x \left\{ \int_{-\infty}^{\infty} F(x, y)\, dG(y) \right\} = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f(x)\, d_x F(x, y) \right\} dG(y) \tag{29}$$

is that the left side of (29) *exist; i.e. that at least one of the quantities* (27) *and* (28) *be finite. This will be true in particular if* $f(x)$ *is bounded from above or from below.*

4. The operation of convolution. An example of the general mixture (25) of distribution functions is the operation of convolution: if $F(x)$, $G(x)$ are two distribution functions in R_1 then $F(x, y) = F(x - y)$ satisfies the conditions of Theorem 4, so that

$$H(x) = \int_{-\infty}^{\infty} F(x - y)\, dG(y) \tag{30}$$

is also a distribution function in R_1, denoted by

$$H(x) = F(x) * G(x). \tag{31}$$

Corresponding to any distribution function $F(x)$ in R_1 is the characteristic function

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x) \tag{32}$$

which in turn uniquely determines $F(x)$ [C, p. 93].

THEOREM 6. *Let* $F(x)$, $G(x)$, $H(x)$ *be distribution functions in* R_1 *and let* $\varphi_1(t)$, $\varphi_2(t)$, $\varphi(t)$ *be the corresponding characteristic functions. Then*

$$H(x) = F(x) * G(x) \tag{33}$$

if and only if

$$\varphi(t) = \varphi_1(t) \cdot \varphi_2(t). \tag{34}$$

PROOF. Assume (33) holds. Since $|e^{itx}| \leq 1$ we have from Theorem 5,

$$
\begin{aligned}
\varphi(t) &= \int_{-\infty}^{\infty} e^{itx}\, d_x\left\{\int_{-\infty}^{\infty} F(x-y)\, dG(y)\right\} \\
&= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} e^{itx}\, d_x F(x-y)\right\} dG(y) \\
&= \int_{-\infty}^{\infty} e^{ity}\left\{\int_{-\infty}^{\infty} e^{it(x-y)}\, d_x F(x-y)\right\} dG(y) \\
&= \int_{-\infty}^{\infty} e^{ity}\left\{\int_{-\infty}^{\infty} e^{itw}\, dF(w)\right\} dG(y) = \varphi_1(t)\cdot\varphi_2(t).
\end{aligned} \tag{35}
$$

The converse implication now follows from the fact that the characteristic function of a distribution determines the latter uniquely.

The importance of the operation $*$ in probability theory arises from the fact that if X, Y are *independent* random variables with respective distribution functions $F(x)$, $G(x)$, and if $Z = X + Y$, then the distribution function $H(x)$ of Z satisfies (33), since for any value of a,

$$
\begin{aligned}
H(a) &= P[X + Y \leq a] = \iint_{x+y\leq a} dF(x)\, dG(y) \\
&= \int_{-\infty}^{\infty}\left\{\int_{x\leq a-y} dF(x)\right\} dG(y) = \int_{-\infty}^{\infty} F(a-y)\, dG(y) = F(a) * G(a),
\end{aligned} \tag{36}
$$

the evaluation of the double integral by an iterated integral following from Fubini's theorem [S, pp. 76–88]. However, (33) may hold without X, Y being independent, and Theorem 6 shows that (34) will then hold also, and conversely.

An example where $H(x) = F(x) * G(x)$ without X, Y being independent has been given by Cramér [C, p. 317, exercise 2]. We shall give another. Let points $0, A, \cdots, F$ in the (x, y)-plane be defined as follows:

$$
O = (0, 0),\ A = (1, 1),\ B = (1/2, 1),\ C = (0, 1/2),\ D = (1, 0),
$$
$$
E = (1, 1/2),\ F = (1/2, 0).
$$

Let $f(x, y)$ have the value 2 inside the quadrilateral $OABC$ and the triangle DEF, and 0 elsewhere. Then if $f(x, y)$ is the joint frequency function of X, Y it is easily seen that X and Y have uniform distributions on the intervals $0 \leq x \leq 1$, $0 \leq y \leq 1$ respectively and that $Z = X + Y$ has the triangular distribution given by (33), although X and Y are not independent.

It would be interesting to know what distribution functions $F(x)$ are such that if $X, Y, Z = X + Y$ are random variables with the distribution functions $F(x)$, $F(x)$, $F(x) * F(x)$ respectively, then X and Y are necessarily independent. A rather trivial example of such a distribution function is the step function $F(x)$ with jumps of $\frac{1}{2}$ at the points $x = 0$ and $x = 1$. It can be shown (oral communication by W. Hoeffding), in generalization of Cramér's example, that no abso-

lutely continuous distribution function (e.g. the normal distribution function) has this property.

5. The problem of random sampling from a mixed population. Let $G(v)$ be a distribution function in the real variable v, and let $F(u, v)$ be for a.e. (relative to the measure corresponding to G) v a distribution function in the real variable u, and for every u a Borel measurable function of v. Let

$$H(u) = \int_{-\infty}^{\infty} F(u, v)\, dG(v); \tag{37}$$

then by Theorem 4 $H(u)$ is a distribution function in R_1. Now define for $x = (x_1, \cdots, x_n)$, $y = (y_1, \cdots, y_n)$

$$\begin{aligned} \bar{H}(x) &= H(x_1) \cdots H(x_n), \\ \bar{G}(y) &= G(y_1) \cdots G(y_n). \end{aligned} \tag{38}$$

Both $\bar{H}(x)$ and $\bar{G}(y)$ are then distribution functions in R_n. In particular, $\bar{H}(x)$ is the distribution function of a random sample of n independent variates each with the distribution function (37). Set

$$\bar{F}(x, y) = F(x_1, y_1) \cdots F(x_n, y_n); \tag{39}$$

then for a. e. (relative to the measure corresponding to $\bar{G}$) y, $\bar{F}(x, y)$ is a distribution function in x, and for every x, $\bar{F}(x, y)$ is a Borel measurable function of y. By Fubini's theorem we have

$$\begin{aligned} \bar{H}(x) &= \int_{-\infty}^{\infty} F(x_1, y_1)\, dG(y_1) \cdot \cdots \cdot \int_{-\infty}^{\infty} F(x_n, y_n)\, dG(y_n) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F(x_1, y_1) \cdots F(x_n, y_n)\, dG(y_1) \cdots dG(y_n) \\ &= \int_{-\infty}^{\infty} \bar{F}(x, y)\, d\bar{G}(y). \end{aligned} \tag{40}$$

Thus $\bar{H}(x)$ is itself a mixture in the sense of Theorem 4. It follows from Theorem 5 that for any Borel measurable function $f(x)$,

$$\int_{-\infty}^{\infty} f(x)\, d\bar{H}(x) = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f(x)\, d_x \bar{F}(x, y) \right\} d\bar{G}(y), \tag{41}$$

if and only if the left side of (41) exists. When written out in full (41) becomes

$$\begin{aligned} &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n)\, d_{x_1} \left\{ \int_{-\infty}^{\infty} F(x_1, y_1)\, dG(y_1) \right\} \\ &\quad \cdots d_{x_n} \left\{ \int_{-\infty}^{\infty} F(x_n, y_n)\, dG(y_n) \right\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} \right. \\ &\quad \left. \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n)\, d_{x_1} F(x_1, y_1) \cdots d_{x_n} F(x_n, y_n) \right\} dG(y_1) \cdots dG(y_n). \end{aligned} \tag{42}$$

Equation (41) is of particular interest in connection with the distribution of a statistic $t = t(x_1, \cdots, x_n) = t(x)$. For any distribution function $J(x)$ let $K(t \mid J)$ denote the distribution function of t when x has the distribution function $J(x)$. If we set

$$f(x) = \begin{cases} 1 & \text{if } t(x) \leq t, \\ 0 & \text{otherwise,} \end{cases} \tag{43}$$

then

$$K(t \mid J) = \int_{-\infty}^{\infty} f(x)\, dJ(x). \tag{44}$$

Hence from (41),

$$\begin{aligned} K(t \mid H(x_1) \cdots H(x_n)) &= K(t \mid \bar{H}) = \int_{-\infty}^{\infty} K(t \mid \bar{F}(x, y))\, d\bar{G}(y) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} K(t \mid F(x_1, y_1) \cdots F(x_n, y_n))\, dG(y_1) \cdots dG(y_n). \end{aligned} \tag{45}$$

As an example, let $t(x)$ be Student's ratio

$$t = n^{\frac{1}{2}} \cdot \bar{x}/s, \tag{46}$$

let

$$F(u, v) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{1}{2}(y-v)^2}\, dy, \tag{47}$$

and let

$$G(v) = \begin{cases} 0 \text{ for } & v < -a, \\ \frac{1}{2} \text{ for } & -a \leq v < a, \\ 1 \text{ for } & a \leq v. \end{cases} \tag{48}$$

Then $H(u)$ will be the distribution function of a mixture in equal proportions of two normal populations with unit variances and with means $-a$, a respectively, and $K(t \mid H(x_1) \cdots H(x_n))$ will be the distribution function of t in random samples of n from this non-normal population. On the other hand, $K(t \mid F(x_1, y_1) \cdots F(x_n, y_n))$ will be the distribution function of t in sampling from successive normal populations with unit variances and means $y_1, \cdots, y_n$ respectively. Relation (45) now becomes

$$K(t \mid H(x_1) \cdots H(x_n)) = \sum_{y_1, \cdots, y_n} K(t \mid F(x_1, y_1) \cdots F(x_n, y_n))/2^n, \tag{49}$$

where the summation is over all 2^n sets $(y_1, \cdots, y_n)$, each y_i being either $-a$ or a. Due to the complexity of $K(t \mid F(x_1, y_1) \cdots F(x_n, y_n))$ (the frequency function of which is discussed in a forthcoming paper by the author), relation

(49) is not very useful. In other cases (45) may afford a considerable simplification in the evaluation of the distribution function of a statistic obtained in random sampling from a mixed population.

REFERENCES

[1] W. Feller, "On the integro-differential equations of purely discontinuous Markoff processes," *Am. Math. Soc. Trans.*, Vol. 48 (1940), p. 488.

[2] R. H. Cameron and W. T. Martin, "An unsymmetric Fubini theorem," *Am. Math. Soc. Bull.*, Vol. 47 (1941), p. 121.

[3] P. R. Halmos, "The decomposition of measures," *Duke Math. Jour.*, Vol. 8 (1941), p. 386.

[4] W. Feller, "On a general class of 'contagious' distributions," *Annals of Math. Stat.*, Vol. 14 (1943), p. 389.

SOME APPLICATIONS OF THE MELLIN TRANSFORM IN STATISTICS

By Benjamin Epstein

Coal Research Laboratory, Carnegie Institute of Technology

1. Summary. It is well known that the Fourier transform is a powerful analytical tool in studying the distribution of sums of independent random variables. In this paper it is pointed out that the Mellin transform is a natural analytical tool to use in studying the distribution of products and quotients of independent random variables. Formulae are given for determining the probability density functions of the product and the quotient $\frac{\xi}{\eta}$, where ξ and η are independent positive random variables with p.d.f.'s $f(x)$ and $g(y)$, in terms of the Mellin transforms $F(s) = \int_0^\infty f(x)\, x^{s-1}\, dx$ and $G(s) = \int_0^\infty g(y) y^{s-1}\, dy$. An extension of the transform technique to random variables which are not everywhere positive is given. A number of examples including Student's t-distribution and Snedecor's F-distribution are worked out by the technique of this paper.

2. Introduction. It is well known [2], [3] that the Fourier transform is a useful analytical tool for studying the distribution of the sums of independent random variables. It is our purpose in this paper to study another transform which is useful in studying the distribution of the product of independent random variables. While it is perfectly true that one can reduce the study of the distribution of the random variable $\xi = \xi_1 \cdot \xi_2 \cdots \xi_n$, the product of n independent random variables $\xi_1, \xi_2, \cdots, \xi_n$, to the study of the distribution of the random variable $\eta = \log \xi = \log \xi_1 + \log \xi_2 + \cdots + \log \xi_n$, the sum of n independent random variables, it seems worth while to study the distribution problem directly. There are advantages inherent in the direct attack on the distribution problem which are lost to a considerable degree, if the problem is so transformed that the Fourier transform becomes applicable. In this paper we shall show that the direct application of the Mellin transform to the study of the distribution of products of independent random variables yields results of interest.

3. Connection between Mellin transforms and products of independent random variables. The key reason for the importance of Fourier transforms in studying the distribution of sums of independent random variables depends on the following result: if ξ_1 and ξ_2 are independent random variables with continuous[1] probability density functions, (henceforth abbreviated as p.d.f.), $f_1(x)$ and $f_2(x)$, respectively, then the p.d.f. $f(x)$ of the random variable $\xi = \xi_1 + \xi_2$ is expressible[1] as

$$f(x) = \int_{-\infty}^{\infty} f_1(x-y) f_2(y)\, dy = \int_{-\infty}^{\infty} f_2(x-y) f_1(y)\, dy. \tag{1}$$

[1] In this paper we shall assume throughout that we are dealing with random variables with continuous p.d.f.'s. The argument can be extended with some changes to distribution functions which are perfectly general, but for simplicity this will not be done here.

But since these expressions are just the Fourier convolutions of $f_1(x)$ and $f_2(x)$, it is small wonder that the Fourier transform plays such a basic role in studying the distribution properties of sums of independent random variables.

Consider now the following result for products of independent random variables (4), (5): if ξ_1 is a random variable with continuous p.d.f. $f_1(x)$ and ξ_2, independent of ξ_1, is a positive random variable with continuous p.d.f. $f_2(x)$, then the p.d.f. $f(x)$ of the random variable $\xi = \xi_1\xi_2$ is expressible[2] as

$$f(x) = \int_0^\infty \frac{1}{y} f_1\left(\frac{x}{y}\right) f_2(y)\, dy. \tag{2}$$

But equation (2) is precisely in the form of a Mellin convolution of $f_1(x)$ and $f_2(x)$ and therefore it may be expected that the Mellin transform should be useful in studying the distribution of products of independent random variables.

It is useful to indicate briefly the properties of the Mellin transform. A detailed treatment of this transform will be found in [6] and we shall, therefore, stress only those portions of the theory of Mellin transforms which are of importance in the field of statistics. By definition, the Mellin transform $F(s)$, corresponding to a function $f(x)$ defined only[3] for $x \geq 0$, is

$$F(s) = \int_0^\infty f(x)x^{s-1}\, dx. \tag{3}$$

Under certain restrictions on $f(x)$ [6, p. 47], $F(s)$ considered as a function of the complex variable s is a function of exponential type, analytic in a strip parallel to the imaginary axis. The width of the strip is governed by the order of magnitude of $f(x)$ in the neighborhood of the origin and for large values of x and, in particular, the strip of analyticity becomes a half-plane if $f(x)$ decays exponentially as $x \to \infty$. There is a reciprocal formula enabling one to go from the transform $F(s)$ to the function $f(x)$. This transformation is:

$$f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^{-s} F(s)\, ds \tag{4}$$

for all x where $f(x)$ is continuous and where the path of integration is any line parallel to the imaginary axis and lying within the strip of analyticity of $F(s)$.

---

[2] More generally [4, p. 411], if ξ_1 and ξ_2 are independent random variables with continuous p.d.f.'s $f_1(x)$ and $f_2(x)$, then the p.d.f. of the random variable $\xi = \xi_1\xi_2$ is expressible as:

$$f(x) = \int_{-\infty}^\infty \frac{1}{|y|} f_1\left(\frac{x}{y}\right) f_2(y)\, dy = \int_{-\infty}^\infty \frac{1}{|y|} f_2\left(\frac{x}{y}\right) f_1(y)\, dy. \tag{2'}$$

In [4] analogous results are given for random variables with perfectly general distribution functions.

[3] The reason for this restriction is that there are technical difficulties in defining a Mellin transform directly for a function defined over $(-\infty, \infty)$. In [6], for instance, the Mellin transform theory is given for functions defined only for positive values of the argument. In statistical terminology this means that we are restricting ourselves for the moment to positive random variables. This is, of course, an unnatural restriction and we shall indicate later in the paper a simple device for treating such questions.

If, in particular, we are interested in applying Mellin transforms to p.d.f.'s of positive[4] random variables, the analysis can be carried out rigorously. Also, as in the case of the Fourier transform, one has the desirable property that there is a one-one correspondence between p.d.f.'s and their transforms.

A number of common p.d.f.'s of positive random variables have simple Mellin transforms. For example see Table 1.

In terms familiar to the mathematical statistician, the Mellin transform of a positive random variable ξ with continuous p.d.f. $f(x)$ is $E(\xi^{s-1})$, where

$$F(s) = E(\xi^{s-1}) = \int_0^\infty x^{s-1} f(x)\, dx. \tag{5}$$

The following three basic properties hold: (*i*) The positive random variable $\eta = a\,\xi$, $a > 0$ has the Mellin transform $G(s) = a^{s-1} F(s)$. This is immediate since

$$G(s) = E(\eta^{s-1}) = E(a^{s-1}\,\xi^{s-1}) = a^{s-1} F(s). \tag{6}$$

(ii) The positive random variable $\eta = \xi^\alpha$ has the Mellin transform $G(s) = F(\alpha s - \alpha + 1)$. To prove this we note that

$$G(s) = E(\eta^{s-1}) = E(\xi^{\alpha s-\alpha}) = F(\alpha s - \alpha + 1). \tag{7}$$

In particular if $\alpha = -1$, i.e., $\eta = \dfrac{1}{\xi}$, then

$$G(s) = F(-s + 2).$$

This is a result which we shall have occasion to use later in the paper.

(iii) If ξ_1 and ξ_2 are independent positive random variables with Mellin transforms $F_1(s)$ and $F_2(s)$, respectively, then the Mellin transform of the product $\eta = \xi_1\xi_2$ is $G(s) = F_1(s)\,F_2(s)$. This is immediate since

$$\begin{aligned} G(s) = E(\eta^{s-1}) = E[(\xi_1\xi_2)^{s-1}] &= E(\xi_1^{s-1})\,E(\xi_2^{s-1}) \\ &= F_1(s)\,F_2(s). \end{aligned} \tag{8}$$

More generally if $\xi_1, \xi_2, \cdots, \xi_n$ are independent positive random variables with Mellin transforms $F_1(s), F_2(s), \cdots, F_n(s)$, then the Mellin transform of the random variable $\eta = \xi_1\xi_2 \cdots \xi_n$ is $G(s) = F_1(s)\,F_2(s) \cdots F_n(s)$. This relationship is fundamental and justifies the introduction of Mellin transforms in studying products of independent random variables.

From (8) it is clear that we can find the p.d.f. $g(y)$ of the random variable η which is the product of two positive independent random variables ξ_1 and ξ_2 with continuous p.d.f.'s $f_1(x)$ and $f_2(x)$. In fact, by the Mellin inversion formula

$$g(y) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} y^{-s} G(s)\, ds = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} y^{-s} F_1(s)F_2(s)\, ds, \tag{9}$$

[4] See footnote 3.

TABLE 1

	p.d.f.	Mellin Transform	Region of Analyticity of Transform
(a)	$f(x) = 1, \quad 0 \leq x \leq 1$ $= 0$, elsewhere	$F(s) = \dfrac{1}{s}$	Half-plane, Re $(s) > 0$
(b)	$f(x) = \dfrac{x^\alpha e^{-x}}{\Gamma(\alpha+1)}, 0 < x < \infty$ $\alpha > -1$	$F(s) = \dfrac{\Gamma(\alpha+s)}{\Gamma(\alpha+1)}$	Half-plane, Re $(s) > -\alpha$
(c)	$f(x) = \dfrac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)} x^\alpha(1-x)^\beta, \qquad 0 < x < 1$ $= 0$, elsewhere $\alpha > -1, \beta > -1$	$F(s) = \dfrac{\Gamma(\alpha+\beta+2)\Gamma(\alpha+s)}{\Gamma(\alpha+\beta+s+1)\Gamma(\alpha+1)}$	Half-plane, Re $(s) > -\alpha$
(d)	$f(x) = \dfrac{\Gamma(\beta)}{\Gamma(\alpha+1)\Gamma(\beta-\alpha-1)} \dfrac{x^\alpha}{(1+x)^\beta}, \quad 0 < x < \infty$ $\alpha > -1, \beta - \alpha > 1$	$F(s) = \dfrac{\Gamma(\alpha+s)\Gamma(\beta-\alpha-s)}{\Gamma(\alpha+1)\Gamma(\beta-\alpha-1)}$	Strip, $-\alpha <$ Re $(s) < \beta - \alpha$

where the path of integration is any line parallel to the imaginary axis and lying within the strip of analyticity of $G(s)$. As in the case of characteristic functions, it can be shown that there is a one-one correspondence between p.d.f.'s and their Mellin transforms. Therefore, it follows that the p.d.f. $g(y)$ computed in this way must be precisely equal to

$$(10) \qquad g(y) = \int_0^\infty \frac{1}{x} f_1\left(\frac{y}{x}\right) f_2(x)\, dx = \int_0^\infty \frac{1}{x} f_2\left(\frac{y}{x}\right) f_1(x)\, dx.$$

It is easy to verify this directly by showing that the Mellin transform of the right-hand side of (10) is $F_1(s)\, F_2(s)$ [6, p. 52], but this will not be done here. The essential point is that Equation (9), (which is sometimes easier to evaluate than Equation (10)), is a consequence of an algebraic formalism which is capable of revealing relationships which would otherwise remain hidden.

The p.d.f. $h(y)$ of $\eta = \frac{\xi_1}{\xi_2}$, the ratio of two positive random variables with continuous p.d.f.'s, can be reduced to finding the p.d.f. of the product of independent random variables ξ_1 and $\frac{1}{\xi_2}$. If $F_1(s)$ and $F_2(s)$ are the Mellin transform corresponding to ξ_1 and ξ_2, respectively, then by (ii) $F_2(-s + 2)$ is the Mellin transform of $\frac{1}{\xi_2}$ and, therefore, the Mellin transform $H(s)$ of $\eta = \frac{\xi_1}{\xi_2}$ is $F_1(s)\, F_2(-s + 2)$. Therefore, the p.d.f. $h(y)$ of η is

$$(11) \quad h(y) = \frac{1}{2\pi i} \int_{c-i,\infty}^{c+i,\infty} y^{-s} H(s)\, ds = \frac{1}{2\pi i} \int_{c-i,\infty}^{c+i,\infty} y^{-s} F_1(s) F_2(-s + 2)\, ds.$$

This formula is useful in finding distributions such as Student's t and Fisher's z.

4. A modified Mellin transform procedure for finding the distribution of the product of independent random variables which are not everywhere positive. Up to this point we have limited ourselves to the application of the Mellin transform to finding the distribution of the product or ratio of two positive independent random variables. While it is true that a number of interesting probability density functions are defined only for positive[5] values of the argument, it is certainly desirable that we be able to treat situations involving random variables capable of taking on both positive and negative values. A simple device for extending the Mellin transform treatment to the more general problem is to decompose the p.d.f.'s $f_1(x)$ and $f_2(x)$ of the independent random variables ξ_1 and ξ_2 into

$$f_1(x) = f_{11}(x) + f_{12}(x),$$
$$f_2(x) = f_{21}(x) + f_{22}(x),$$

[5] For example, distributions of type 3, the χ^2 distribution, the distribution of the sample standard deviation and sample variance, the distribution of an even power of a random variable, etc. are all defined only for positive values of the argument.

where[6]

$$f_{11}(x) = 0, x < 0, \qquad f_{12}(x) = 0, x > 0,$$

$$f_{21}(x) = 0, x < 0, \qquad f_{22}(x) = 0, x > 0,$$

and then to operate on the pairs $[f_{11}(x), f_{21}(x)]$, $[f_{11}(x), f_{22}(x)]$, $[f_{12}(x), f_{21}(x)]$, and $[f_{12}(x), f_{22}(x)]$ separately. More specifically, the frequency distribution $h(y)$ corresponding to the random variable $\eta = \xi_1\xi_2$ is made up of the sum of four components $h_1(y)$, $h_2(y)$, $h_3(y)$, and $h_4(y)$. To compute $h_1(y)$ one can apply the Mellin transform directly to the evaluation of the expression

$$h_1(y) = \int_0^\infty \frac{1}{x} f_{11}\left(\frac{y}{x}\right) f_{21}(x)\, dx,$$

since both $f_{11}(x)$ and $f_{21}(x)$ are zero for negative values of x. The function $h_1(y)$ is zero for $y < 0$. To compute $h_2(y)$ we first evaluate

$$h_2^*(y) = \int_0^\infty \frac{1}{x} f_{11}\left(\frac{y}{x}\right) f_{22}(-x)\, dx.$$

Again $f_{11}(x)$ and $f_{22}(-x)$ are zero for negative values of x and, therefore, the conventional Mellin transform can be applied in determining $h_2^*(y)$. It is clear that $h_2^*(y) = 0$ for $y < 0$ and, therefore, $h_2(y) = h_2^*(-y) = 0$ for $y > 0$. Similarly, one can find $h_3(y)$ and $h_4(y)$ where $h_3(y) = 0$ for $y > 0$ and $h_4(y) = 0$ for $y < 0$, and it is readily seen that[7]

$$h(y) = h_1(y) + h_2(y) + h_3(y) + h_4(y)$$

is the desired p.d.f. of $\eta = \xi_1\xi_2$.

5. Examples of use of Mellin transforms in evaluating the product and quotient of independent random variables. Example 1: *The distribution of* $\eta = \xi_1\xi_2$, *where* ξ_1 *and* ξ_2 *are independent random variables with p.d.f.'s* $f_1(x)$ *and* $f_2(x)$, *respectively, where*

$$f_1(x) = f_2(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty.$$

In this case

$$f_1(x) = f_{11}(x) + f_{12}(x),$$

and

$$f_2(x) = f_{21}(x) + f_{22}(x),$$

[6] Of course, f_{11}, f_{12}, f_{21}, and f_{22} are generally not p.d.f.'s since $\int_0^\infty f_{11}(x)\, dx$, $\int_{-\infty}^0 f_{12}(x)\, dx$, $\int_0^\infty f_{21}(x)\, dx$, $\int_{-\infty}^0 f_{22}(x)\, dx$ are no longer necessarily equal to one.

[7] As in footnote 6, h_1, h_2, h_3, and h_4 are, in general, not p.d.f.'s.

where

$$f_{11}(x) = 0, x < 0; f_{12}(x) = 0, x > 0;$$

$$f_{21}(x) = 0, x < 0; f_{22}(x) = 0, x > 0.$$

The random variable $\eta = \xi_1\xi_2$ has a p.d.f. $h(y) = h_1(y) + h_2(y) + h_3(y) + h_4(y)$ where

$$\begin{aligned} &h_1(y) \text{ is associated with } [f_{11}(x), f_{21}(x)], \\ &h_2(y) \text{ is associated with } [f_{11}(x), f_{22}(x)], \\ &h_3(y) \text{ is associated with } [f_{12}(x), f_{21}(x)], \end{aligned}$$

and $h_4(y)$ is associated with $[f_{12}(x), f_{22}(x)]$.

It is sufficient to evaluate

$$\begin{aligned} h_1(y) &= \int_0^\infty \frac{1}{x} f_{11}\left(\frac{y}{x}\right) f_{21}(x)\, dx. \\ &= \int_0^\infty \frac{1}{x} f_{21}\left(\frac{y}{x}\right) f_{11}(x)\, dx. \end{aligned}$$

In this case

$$F_{11}(s) = \int_0^\infty x^{s-1} f_{11}(x)\, dx = \int_0^\infty x^{s-1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \frac{2^{\frac{1}{2}(s-3)}}{\sqrt{\pi}} \Gamma(s/2),$$

analytic for $\text{Re}(s) > 0$
and

$$F_{21}(s) = \int_0^\infty x^{s-1} f_{21}(x)\, dx = \frac{2^{\frac{1}{2}(s-3)}}{\sqrt{\pi}} \Gamma(s/2).$$

Therefore,

$$H_1(s) = F_{11}(s)F_{21}(s) = \frac{2^{s-3}}{\pi} \Gamma^2(s/2)$$

$$\begin{aligned} h_1(y) &= \frac{1}{2\pi i} \int_{c-i,\infty}^{c+i,\infty} y^{-s} H_1(s)\, ds \\ &= \frac{1}{2\pi i} \int_{c-i,\infty}^{c+i,\infty} y^{-s} \frac{2^{s-3}}{\pi} \Gamma^2(s/2)\, ds, \qquad c > 0 \\ &= \frac{1}{2\pi} K_0(y), \qquad y > 0 \quad [6, \text{p. } 197] \end{aligned}$$

where $K_0(y)$ is Bessel's function of the second kind with a purely imaginary argument of zero order. Similarly

$$h_2(y) = \frac{1}{2\pi} K_0(y), \qquad y < 0$$

$$h_3(y) = \frac{1}{2\pi} K_0(y), \qquad y < 0$$

$$h_4(y) = \frac{1}{2\pi} K_0(y), \qquad y > 0.$$

Therefore, $h(y) = h_1(y) + h_2(y) + h_3(y) + h_4(y)$

$$= \frac{1}{\pi} K_0(y), \qquad -\infty < y < \infty,$$

and this is the desired p.d.f. This result has been found by other methods and is given in [1, p. 1].

Example 2: *The distribution of* $\eta = \frac{\xi_1}{\xi_2}$ *where* ξ_1 *and* ξ_2 *are independent random variables with p.d.f.'s* $f_1(x)$ *and* $f_2(x)$, *respectively, where*

$$f_1(x) = f_2(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < y < \infty.$$

As in Example 1, one splits the determination of $h(y)$, the p.d.f. of η, into four parts: $h_1(y)$, $h_2(y)$, $h_3(y)$, $h_4(y)$. In the notation of Example 1 it is easy to show that $H_{11}(s)$ the Mellin transform of $h_1(y)$ is

$$F_{11}(s)F_{21}(-s+2) = \frac{2^{\frac{1}{2}(s-3)}}{\sqrt{\pi}} \Gamma(s/2) \frac{2^{\frac{1}{2}(s-3)}}{\sqrt{\pi}} \Gamma(-s/2+1) = \frac{1}{4} \frac{1}{\sin \frac{s\pi}{2}};$$

$$h_1(y) = \frac{1}{2\pi i} \int_{c-i,\infty}^{c+i,\infty} y^{-s} H(s)\, ds, \qquad 0 < c < 2,$$

$$= \frac{1}{2\pi i} \int_{c-i,\infty}^{c+i,\infty} \frac{1}{4} \frac{y^{-s}\, ds}{\sin \frac{s\pi}{2}}$$

$$= \frac{1}{2\pi} \frac{1}{1+y^2}, \qquad y \geq 0.$$

Similarly

$$h_2(y) = \frac{1}{2\pi} \frac{1}{1+y^2}, \qquad y \leq 0,$$

$$h_3(y) = \frac{1}{2\pi} \frac{1}{1+y^2}, \qquad y \leq 0,$$

$$h_4(y) = \frac{1}{2\pi} \frac{1}{1+y^2}, \qquad y \geq 0.$$

Therefore, $h(y) = h_1(y) + h_2(y) + h_3(y) + h_4(y)$

$$= \frac{1}{\pi} \frac{1}{1+y^2}, \qquad -\infty < y < \infty.$$

This result has been found by other methods and given in [4, p. 411].

Example 3: *F-Distribution.* Let $\xi_1, \cdots, \xi_m, \eta_1, \cdots, \eta_n$ be $(m+n)$ independ-

ent random variables, each normally distributed with mean zero and standard deviation σ. Let

$$\xi = \sum_{i=1}^{m} \xi_i^2, \qquad \eta = \sum_{j=1}^{n} \eta_j^2.$$

We want to find the p.d.f. $h(z)$ of ζ where $\zeta = \xi/\eta$. The p.d.f.'s $f(x)$ and $g(y)$ of ξ and η, respectively, are:

$$f(x) = \frac{x^{m/2-1}\, e^{-x/2\sigma^2}}{2^{m/2}\sigma^m\Gamma(m/2)}, \qquad x > 0,$$

and

$$g(y) = \frac{y^{n/2-1}\, e^{-y/2\sigma^2}}{2^{n/2}\,\sigma^n\,\Gamma(n/2)}, \qquad y > 0.$$

In this case

$$F(s) = \frac{2^{s-1}\sigma^{2s-2}\Gamma\left(s + \frac{m}{2} - 1\right)}{\Gamma(m/2)}, \qquad \text{analytic for Re } (s) > 1 - \frac{m}{2},$$

and

$$G(s) = \frac{2^{s-1}\sigma^{2s-2}\Gamma\left(s + \frac{n}{2} - 1\right)}{\Gamma(n/2)}, \qquad \text{analytic for Re } (s) > 1 - \frac{n}{2}.$$

The p.d.f. $h(z)$ has Mellin transform

$$H(s) = F(s)\, G(-s + 2)$$
$$= \frac{\Gamma\left(s + \frac{m}{2} - 1\right)\Gamma\left(-s + \frac{n}{2} + 1\right)}{\Gamma(m/2)\Gamma(n/2)}.$$

Therefore,

$$h(z) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} z^{-s} H(s)\, dz, \qquad -\frac{m}{2} + 1 < c < \frac{n}{2} + 1,$$
$$= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma(m/2)\Gamma(n/2)}\; \frac{z^{m/2-1}}{(z+1)^{\frac{1}{2}(m+n)}}, \qquad z > 0.$$

A convenient way of carrying out the inversion is to use formula (d) in Table 1.

In a similar way one can find Student's distribution, i.e., the distribution of $\zeta = \xi_0/\eta$, where $\eta = \sqrt{\sum_1^n \xi_\nu^2/n}$, and where $\xi_0, \xi_1, \cdots, \xi_n$ are $n + 1$ independent random variables each having the distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2/2\sigma^2}, \qquad -\infty < x < \infty.$$

It should be mentioned in conclusion that the Mellin transform is a natural tool to use in situations involving the products and quotients of independent uniformly distributed random variables, or in finding products and/or quotients and/or Beta-distribution. In such cases formulae (b), (c) and (d) in Table 1 are useful.

REFERENCES

[1] C. C. CRAIG, "On the frequency function of xy," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 1–15.

[2] H. CRAMÉR, *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics, No. 36, Cambridge, 1937.

[3] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[4] J. H. CURTISS, "On the distribution of the quotient of two chance variables," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 409–421.

[5] E. V. HUNTINGTON, "Frequency distribution of product and sum," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 195–198.

[6] E. C. TITCHMARSH, *Introduction to the Theory of Fourier Integrals*, Clarendon Press, Oxford, 1937.

THE ESTIMATION OF LINEAR TRENDS

By G. W. Housner and J. F. Brennan

California Institute of Technology

1. Summary. This paper deals with the problem of bivariate regression where both variates are random variables having a finite number of means distributed along a straight line. A regression statistic is derived which is independent of change in scale so that a prior knowledge of the frequency distribution parameters is not required in order to obtain a unique estimate. The statistic is shown to be consistent. The efficiency of the estimate is discussed and its asymptotic distribution is derived for the case when the random variables are normally distributed. A numerical example is presented which compares the performance of the statistic of this paper with that of other commonly used statistics. In the example it is found that the method of estimation proposed in this paper is more efficient.

2. Introduction. A problem that often arises in statistical work is the estimation of linear trends. In the general problem it is known or presumed that a linear functional relation exists among a set of variables of the form,

$$a + b_1X + b_2Y + b_3Z \cdots = 0.$$

The observed values of the variables are of the form

$$x_{ik} = X_i + \epsilon_{ik}, \qquad y_{ik} = Y_i + \eta_{ik}, \quad \text{etc.}$$

That is, the x_{ik} are random variables with means X_i and $k = 1, 2, \cdots N_i$ observed values of x are associated with the mean X_i. The ordering of the X_i is according to magnitude. Similarly there are the observed values y_{ik}, z_{ik} and so forth. The ϵ_{ik} are random variables, with the same distribution for all i, with zero means. On the basis of a sample $O_n(x_{ik}, y_{ik}, z_{ik}, \cdots)$ it is desired to estimate the coefficients $a, b_1, b_2, b_3, \cdots$. One method used to estimate the coefficients is that of "weighted regression" which is essentially an application of the method of least squares. The problem has been studied by R. Allen, A. Wald and others.[1] The chief difficulty has been that the proposed methods of estimation require an a priori knowledge of the variances of the random variables. Wald has proposed a statistic which avoids this difficulty but which may have a relatively low efficiency in cases often encountered in practice. In this paper there is described a bivariate statistic which appears to have comparatively high precision and which does not require prior knowledge of the variances of the random variables. A numerical example is given at the end of the paper to illustrate the comparative performances of different methods of estimation.

[1] For a brief history of work done on this problem see the paper by A. Wald in the *Annals of Math. Stat.*, Vol. 11 (1940), p. 284.

3. The Regression statistic. In the case of the bivariate problem, consider a sample

$$O_n(x_{ik}, y_{ik}), i = 1, 2, \cdots, n$$

and

$$k = 1, 2, \cdots, N_i,$$

where N_i sample values x_i, y_i are distributed about mean X_i, Y_i. Let the means be related by $Y_i = a + bX_i$ and let the random variables x_i be independent and have the same frequency distribution with variance σ_x^2 for all i and the random variables y_i have independent frequency distributions with variance σ_y^2 the same for all i. An appropriate statistic for estimating b is obtained by noting that a pair of sample points (x_{ik}, y_{ik}), (x_{jl}, y_{jl}) gives a sample value of the change in y corresponding to a change in x. It may thus be said that a sample value of b is

$$\hat{b}_{ik,jl} = \frac{y_{ik} - y_{jl}}{x_{ik} - x_{jl}}. \tag{1}$$

Making use of the fact that

$$y_{ik} = a + bx_{ik} + \eta_{ik} - b\epsilon_{ik} \tag{2}$$

equation (1) may be written

$$(x_{ik} - x_{jl})\, \hat{b}_{ik,jl} = (x_{ik} - x_{jl})\, b + (\eta_{ik} - \eta_{jl}) - b(\epsilon_{ik} - \epsilon_{jl}).$$

Summing this equation over all combinations of points there is obtained

$$b = \frac{\sum_i \sum_j \sum_k \sum_l (y_{ik} - y_{jl})}{\sum_i \sum_j \sum_k \sum_l (x_{ik} - x_{jl})} - \frac{\sum_i \sum_j \sum_k \sum_l ((\eta_{ik} - \eta_{jl}) - b(\epsilon_{ik} - \epsilon_{jl}))}{\sum_i \sum_j \sum_k \sum_l (x_{ik} - x_{jl})}. \tag{3}$$

The summations in the above expression are to be carried out for

$$l = 1, 2, \cdots, N_j; k = 1, 2, \cdots, N_i; j = 1, 2, \cdots, (i - 1); i = 1, 2, \cdots, n.$$

The first term on the right side of equation (3) is an estimate of b and the second term represents the deviation of the estimate from the true value. Accordingly, we take as an estimate of b the statistic

$$\hat{b} = \frac{\sum_i \sum_j \sum_k \sum_l (y_{ik} - y_{jl})}{\sum_i \sum_j \sum_k \sum_l (x_{ik} - x_{jl})}. \tag{4}$$

This requires, of course, that the denominator be not equal to zero. Summing out the subscripts k and l reduces (4) to

$$\hat{b} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{i-1} N_i N_j(\bar{y}_i - \bar{y}_j)}{\sum_{i=1}^{n} \sum_{j=1}^{i-1} N_i N_j(\bar{x}_i - \bar{x}_j)}$$

where $\bar{y}_i$ is the mean value of the y_{ik} and so forth. Summing out the subscript j gives

$$\hat{b} = \frac{\sum_i \left(N_i \bar{y}_i \sum_1^{i-1} N_j - N_i \sum_1^{i-1} N_j \bar{y}_j \right)}{\sum_i \left(N_i \bar{x}_i \sum_1^{i-1} N_j - N_i \sum_1^{i-1} N_j \bar{x}_j \right)}. \tag{5}$$

This expression may be put in a more convenient form by using the identity

$$\sum_{i=1}^n \left(N_i \sum_1^{i-1} N_j \bar{y}_j \right) = \sum_{i=1}^n \left(N_i \bar{y}_i \sum_{i+1}^n N_j \right) = \sum_{i=1}^n \left(N_i \bar{y}_i \left(\sum_1^n N_j - \sum_1^i N_j \right) \right).$$

With this substitution equation (5) becomes

$$\hat{b} = \frac{\sum_{i=1}^n \left[N_i \bar{y}_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) \right]}{\sum_{i=1}^n \left[N_i \bar{x}_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) \right]}. \tag{6}$$

This is the statistic for estimating the linear trend of bivariate data. It may be noted that its derivation is not based on the notion of fitting a line to the sample points. A line $y = \hat{a} + \hat{b}x$ may be fitted to the sample points by making it pass through the mean of the sample points, that is, by using the following estimate:

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

where $\bar{y}$ and $\bar{x}$ are the means of all the y_{ik} and x_{ik} respectively.

4. Consistency of the estimate. Having established the statistics $\hat{b}$ and $\hat{a}$ it is desirable to examine the consistency and efficiency of the estimates, particularly for $\hat{b}$. To determine that $\hat{b}$ is a consistent estimate we investigate the behavior of (6) as the number of sample points increases, that is, as the $N_i \to \infty$. We wish first to establish the following identity. Consider the sum of the following array of terms:

$$\begin{gathered} N_1(N_1 + N_2 + \cdots + N_n) \\ N_2(N_1 + N_2 + \cdots + N_n) \\ \vdots \\ N_n(N_1 + N_2 + \cdots + N_n) \end{gathered}$$

The sum may be written $\sum_1^n N_i \sum_1^n N_j$. Since the array is skew symmetrical the expression $2 \sum_1^n N_i \sum_1^i N_j$ also gives the sum of the array except for the fact that the terms along the principal diagonal are counted twice. We have, therefore

$$\sum_1^n N_i \sum_1^n N_j = 2 \sum_1^n N_i \sum_1^i N_j - \sum_1^n N_i^2.$$

Rearranging terms we obtain the identity

$$\sum_1^n \left[N_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) \right] = 0. \tag{7}$$

Now substituting (2) into (6) and making use of (7) there is obtained,

$$\hat{b} = b + \frac{\sum_1^n \left[N_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) (\bar{\eta}_i - b\bar{\epsilon}_i) \right]}{\sum_1^n \left[N_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) \bar{x}_i \right]}. \tag{8}$$

The $\bar{\eta}_i$ and $\bar{\epsilon}_i$ are random variables with zero means so that as $N_i \to \infty$ the sample means $\bar{\eta}_i$ and $\bar{\epsilon}_i$ converge in probability to zero. As $N_i \to \infty$, $\bar{x}_i$ converges in probability to its mean X_i. In view of (7) and that the denominator in (8) is not equal to zero the last term in (8) converges in probability to zero and $\hat{b} \to b$. The estimate is therefore consistant. A similar argument also shows the estimate $\hat{a}$ to be consistent.

5. Efficiency of the estimate. A general investigation of the efficiency of the estimate $\hat{b}$ is beyond the scope of this paper. We may note, however, that the efficiency of the estimate can be made to depend upon the grouping of the data, that is, the optimum efficiency of the estimate may depend upon the omission of some of the pairs $(y_{ik} - y_{jl})$ from the estimate. The maximum efficiency is obtained for $\hat{b}$ when the second term in (3) is minimized. This requires prior knowledge of the frequency distribution of the random variables x and y; however, in applications a recognition of (3) may often indicate a practical method of increasing the efficiency.

In what follows we make an investigation of the precision of the estimate $\hat{b}$ for a special case which is of some practical interest. Let x and y be random variables as defined in the first part of the paper and consider the new variables defined by $\hat{b} = \frac{v}{u}$ that is,

$$u = \sum_1^n \left[N_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) \bar{x}_i \right]$$

$$v = \sum_1^n \left[N_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) \bar{y}_i \right].$$

The random variables u and v are then independently distributed with joint probability element $f(u)\,f(v)\,du\,dv$. Making the change of variable $u = r\cos\theta$, $v = r\sin\theta$ the probability element becomes $f(r, \theta)dr\,d\theta$ where $\tan\theta = u$. Integrating out the variable r gives the probability element for θ. In what follows we investigate the distribution of θ for the case where x and y are normally distributed with the same variance. Since u and v are linear functions of x and y respectively they are also normally distributed with the same standard deviation.

We designate the means of u and v by m_1 and m_2 respectively and the standard deviation by σ. The probability element in u and v is then

$$\frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}[(u-m_1)^2+(v-m_2)^2]\right\} du\, dv. \tag{9}$$

Changing variables to r, θ and setting $m_1 = \bar{r}\cos\bar{\theta}$, $m_2 = \bar{r}\sin\bar{\theta}$ we obtain the following probability element:

$$\frac{r}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}[(r\cos\theta-\bar{r}\cos\bar{\theta})^2+(r\sin\theta-\bar{r}\sin\bar{\theta})]\right\} dr\, d\theta.$$

Completing the square in r and substituting $\phi = \theta - \bar{\theta}$ there is obtained

$$\frac{r}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}(r-\bar{r}\cos\phi)^2\right\} \exp -\tfrac{1}{2}\left(\frac{\bar{r}\sin\phi}{\sigma}\right)^2\Big\} dr\, d\phi. \tag{10}$$

To integrate out r make further change of variable

$$t = \frac{r}{\sigma} - \frac{\bar{r}}{\sigma}\cos\phi.$$

Setting $\frac{\bar{r}}{\sigma}\cos\phi = w$ for convenience in notation there is obtained

$$\left(\frac{1}{2\pi}t\exp\left\{-\frac{t^2}{2}\right\} + \frac{w}{2\pi}\exp\left\{-\frac{t^2}{2}\right\}\right)\exp\left\{-\tfrac{1}{2}\left(\frac{\bar{r}^2}{\sigma^2}-w^2\right)\right\} dt\, d\phi.$$

The variable t is to be integrated out of this expression. The corresponding limits of integration are exhibited by

$$\begin{aligned}\frac{w}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\bar{r}^2}{\sigma^2}-w^2\right)\right\}&\left(\frac{1}{\sqrt{2\pi}}\int_{-w}^{+\infty}\frac{t}{w}\exp\left\{-\frac{t^2}{2}\right\}dt\right.\\ &\left.+\frac{1}{\sqrt{2\pi}}\int_{-w}^{+\infty}\exp\left\{-\frac{t^2}{2}\right\}dt\right)d\phi.\end{aligned} \tag{12}$$

Now as the number of points in the original sample increases the value of $\bar{r}$ also increases and as $\frac{\sigma}{\bar{r}} \to 0$, with $|\phi| < \frac{\pi}{2}$, the value of $w \to \infty$. In this case then (12) approaches asymptotically to

$$\frac{\bar{r}}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{\sin\phi^2}{\sigma/\bar{r}}\right)^2\right\} d(\sin\phi).$$

As $\sigma/\bar{r} \to 0$ this distribution shows that ϕ converges in probability to zero and that the distribution approaches asymptotically to the normal form

$$\frac{1}{\sqrt{2\pi}\sigma/\bar{r}}\exp\left\{-\tfrac{1}{2}\left(\frac{\phi}{\sigma/\bar{r}}\right)^2\right\} d\phi. \tag{13}$$

It is required then to examine the conditions under which $\sigma/\bar{r}$ assumes small values. If the variance of the original variables x_i and y_i is designated by σ_1^2

then since u and v are linear functions of x_i and y_i respectively the variance of u and of v is

$$(14) \qquad \sigma^2 = \sigma_1^2 \sum_1^n \left\{ \left[N_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) \right]^2 \left(\frac{1}{N_i} \right) \right\}.$$

Now $\bar{r}^2$ is the sum of the squares of the means of u and v so that

$$(15) \qquad \bar{r}^2 = (1 + b^2) \left\{ \sum_1^n \left[N_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) X_i \right] \right\}^2.$$

Dividing (14) by (15) we obtain

$$(16) \qquad \left(\frac{\sigma}{\bar{r}} \right)^2 = \frac{\sigma_1^2}{1 + b^2} \frac{\sum_1^n \left\{ \left[N_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) \right]^2 \left(\frac{1}{N_i} \right) \right\}}{\left\{ \sum_1^n \left[N_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) X_i \right] \right\}^2}.$$

Inspection of (16) indicates that as the number of sample points N_i increases the value of $\left(\frac{\sigma}{\bar{r}} \right)^2$ decreases rapidly. To illustrate this we examine some particular cases. Consider first the case of four equally spaced means $X_i = 3i\sigma_1$, $(i = 1, 2, 3, 4)$ and let there be one sample point for each mean $(N_i = 1)$. With these values there is obtained,

$$\left(\frac{\sigma}{\bar{r}} \right)^2 = \frac{0.022}{1 + b^2}.$$

For $b = 1$ the range $-9° < \phi < +9°$ includes 95% of the population defined by (13). As the number of points N_i is increased or as the number of means X_i is increased the value of $\left(\frac{\sigma}{\bar{r}} \right)^2$ decreases rapidly. Consider now eight equally spaced means $X_i = 3_i\sigma_1$, $(i = 1, 2, \cdots, 8)$ with again one sample point for each mean $(N_i = 1)$. With these values there is obtained

$$\left(\frac{\sigma}{\bar{r}} \right)^2 = \frac{0.00045}{1 + b^2}.$$

For $b = 1$ the range $-1° < \phi < +1°$ includes 95% of the population defined by (13).

It is clear that a very high degree of precision is obtained with the estimate $\hat{b}$ when there is a considerable number of sample points. However, this will also be true in general of other statistics and it is really of interest to compare precisions in those cases where the statistics have a relatively low precision. A detailed comparison is beyond the scope of this paper. However, a direct comparison can be made very easily in the particular case when x_i is a fixed variate

and only y_i is a random variable. For the sake of simplicity, let each $N_i = 1$ then the statistic for estimating b is

$$(17) \qquad \hat{b} = \frac{\sum_1^n i(y_i - \bar{y})}{\sum_1^n i(x_i - \bar{x})} = \frac{\sum_1^n y_i(i - \bar{\imath})}{\sum_1^n x_i(i - \bar{\imath})}.$$

Since $\hat{b}$ is a linear function of the y_i by a well known theorem its variance is

$$(18) \qquad \sigma_{\hat{b}}^2 = \sigma_y^2 \sum_1^n \left(\frac{i - \bar{\imath}}{\sum_1^n x_i(i - \bar{\imath})} \right)^2.$$

The customary least squares regression line of y on x gives for the estimate of b and its variance

$$\hat{b}_R = \frac{\sum_1^n y_i(x_i - \bar{x})}{\sum_1^n x_i(x_i - \bar{x})} \qquad \sigma_{\hat{b}_R}^2 = \sigma_y^2 \sum_1^n \left(\frac{x_i - \bar{x}}{\sum_1^n x_i(x_i - \bar{x})} \right)^2.$$

In the particular case when the x_i are equally spaced, $x_i = ci + d$, the estimates $\hat{b}$ and $\hat{b}_R$ are identical:

$$(19) \qquad \hat{b} = \hat{b}_R = \frac{12}{cn(n^2 - 1)} \sum_1^n y_i(i - \bar{\imath}).$$

6. Numerical example. From a practical point of view the case where x and y are random variables is of greater interest than where x is a fixed variate. We give a numerical example of this case comparing the statistic $\hat{b}$ with several other statistics. Consider the case where there is one sample point for each mean X_i. We shall evaluate the following:

1). The statistic of this paper which for this case is

$$\hat{b}_1 = \frac{\sum_1^n y_i(i - \bar{\imath})}{\sum_1^n x_i(i - \bar{\imath})}.$$

2). The statistic obtained by minimizing the sum of the squares of the y deviations only

$$\hat{b}_2 = \frac{\sum_1^n y_i(x_i - \bar{x})}{\sum_1^n x_i(x_i - \bar{x})}.$$

3). The statistic obtained by minimizing the sum of the squares of the orthogonal deviations

$$\hat{b}_3 = \frac{\sum_1^n (y_i - \bar{y})^2 - \sum_1^n (x_i - \bar{x})^2 + \left[n \sum_1^n (y_i - \bar{y})^2 - n \sum_1^n (x_i - \bar{x})^2 + 4\left(\sum_1^n (y_i - \bar{y})(x_i - \bar{x})\right)^2 \right]^{\frac{1}{2}}}{\sum_1^n (y - \bar{y})(x - \bar{x})}$$

TABLE I

Set	x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
1	1.1	1.4	2.4	2.0	3.0	2.7	3.6	4.3
2	1.2	1.4	2.2	2.0	3.4	3.1	3.8	4.2
3	1.0	1.4	1.6	2.1	2.8	3.2	4.4	4.3
4	0.6	0.7	1.8	2.0	3.3	2.6	3.8	4.0
5	0.7	1.4	1.7	1.7	2.7	3.4	4.1	4.1
6	1.0	1.2	1.6	2.1	2.9	2.6	3.6	4.0
7	1.3	0.7	1.7	2.1	2.7	2.9	4.0	3.6

TABLE II

Set	b_1	b_2	b_3	b_4
1	1.160	1.068	1.220	1.162
2	1.056	1.009	1.059	1.027
3	0.860	0.843	0.803	0.870
4	0.946	0.896	0.924	0.830
5	0.875	0.867	0.913	1.000
6	0.978	0.939	0.981	0.846
7	1.044	0.959	1.045	1.000
Mean.	0.990	0.940	0.996	0.962
7 × Sample Var.	0.0686	0.0373	0.1058	0.0834

4). The statistic proposed by Wald[2]

$$\hat{b}_4 = \frac{\sum_1^{n/2} y_i - \sum_{n/2}^n y_i}{\sum_1^{n/2} x_i - \sum_{n/2}^n x_i}.$$

We apply these statistics to sample data having four means $X_i = i$ and $Y_i = i, (i = 1, 2, 3, 4)$. By means of a table of random numbers seven sets of data were

[2] Loc. cit.

obtained, each set having one sample point corresponding to each mean. These sample points are described by Table I where it will be noted that the sample points were drawn from a discrete distribution. The estimates obtained from the four statistics are exhibited in Table II.

If the 28 sample points are treated as a single set of data and the four statistics in their appropriate forms are applied, there is obtained the following set of estimates:

$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$
0.9768	0.9183	0.9786	0.9496.

The preceding computations show that the estimate $\hat{b}_2$ is inferior to the other estimates, as would be expected. The estimate $\hat{b}_3$ is most accurate when the 28 sample points are treated as a single set of data with the estimate $\hat{b}_1$ being only very slightly less accurate, $\hat{b}_1 = 0.9768$ as compared to $\hat{b}_3 = 0.9786$. When the individual sets of sample points 1 to 7 are considered it is seen that the estimate $\hat{b}_1$ is most accurate with the estimate $\hat{b}_3$ rather less accurate; the estimate $\hat{b}_1$ is more precise than $\hat{b}_3$, the sample variances being in the ratio $0.0686 \div 0.1058 = 0.65$. From a practical viewpoint we may also point out that the computation of $\hat{b}_1$ requires very much less labor than the computation of $\hat{b}_3$.

ON THE EFFECT OF DECIMAL CORRECTIONS ON ERRORS OF OBSERVATION

By Philip Hartman and Aurel Wintner

The Johns Hopkins University

1. Summary. Let t be the true value of what is being measured and suppose that the error of observation is a symmetric normal distribution of standard deviation σ. The "rounding-off" error due to the reading of measurements to the nearest unit has a distribution and an expected value depending on t and σ. It is shown that, for a fixed $\sigma > 0$, the expected value of the decimal correction, $r(t; \sigma)$, is an analytic function of t which is odd, of period 1, positive for $0 < t < \frac{1}{2}$, and has a convex arch as its graph on $0 \leqq t \leqq \frac{1}{2}$. Furthermore, if $0 < t < \frac{1}{2}$, both $r(t; \sigma)$ and its maximum value, $\operatorname{Max}_t r(t; \sigma)$, are decreasing functions of σ.

2. Introduction. Let X be an error of observation and let $\phi(x)$ denote the density of probability of the distribution of X. In particular,

$$(1) \qquad \int_{-\infty}^{+\infty} \phi(x)\, dx = 1, \qquad \text{where } \phi(x) \geqq 0.$$

If t is any fixed number, the density of probability of the distribution of $X + t$ is $\phi(x - t)$.

Besides the "instrumental error of observation", X, there is another error, that of the "rounding-off", which is carried along in the registration of the measurements. It is introduced by the circumstance that, if $\cdots$, b, a are digits, and if b denotes the last digit considered, then decimal fractions such as $\cdots ba$ and $\cdots ba \cdots$ are registered as $\cdots b$ if $a < 5$ and as $\cdots (b + 1)$ if $a > 5$. Let the unit, in which the measurements are expressed, be so chosen that the first digit neglected becomes the first digit following the decimal point, i.e., that the error of the "rounding-off" is between $\pm\frac{1}{2}$. Then, if t denotes the *true* value of what is being measured, the remark made after (1) shows that *the probability that the error of the decimal corrections be less than* x is given by

$$\sum_{n=-\infty}^{\infty} \int_{n-\frac{1}{2}}^{n-\frac{1}{2}+x} \phi(u - t)\, du,$$

if $|x| \leqq \frac{1}{2}$, whereas this probability is 0 or 1 according as $x < -\frac{1}{2}$ or $x > \frac{1}{2}$. Since the last series can be written in the form

$$(2) \qquad \sum_{n=-\infty}^{\infty} \int_{-\frac{1}{2}}^{-\frac{1}{2}+x} \phi(u + n - t)\, du = \int_{-\frac{1}{2}}^{-\frac{1}{2}+x} \sum_{n=-\infty}^{\infty} \phi(u + n - t)\, du, \qquad (\phi \geqq 0),$$

it follows that the density of probability of the error due to the decimal corrections is

$$(3) \qquad \sum_{n=-\infty}^{\infty} \phi(x + n - t) \text{ if } |x| < \tfrac{1}{2}, \text{ and } 0 \text{ if } |x| > \tfrac{1}{2}.$$

Consequently, *if $r = r(t)$ denotes the expected value of the decimal error induced on the "true" value, t, of the observations, then*

$$r(t) = \int_{|x|<\frac{1}{2}} x \sum_{n=-\infty}^{\infty} \phi(x + n - t)\, dx. \tag{4}$$

Formula (4) is known[1]. It is usually based on its intuitive interpretation which results if, on the one hand, (4) is written in the form

$$r(t) = \int_{-\infty}^{\infty} s(x)\phi(x - t)\, dx, \tag{5}$$

where

$$s(x) = x \text{ if } -\tfrac{1}{2} < x < \tfrac{1}{2} \text{ and } s(x) = s(x + 1), \qquad -\infty < x < \infty, \tag{6}$$

and, on the other hand, the periodic function (6) is thought of as representing the *uniform* distribution of the error of "rounding-off" over the *arithmetical continuum* over a period,

$$|x - n| < \tfrac{1}{2}, \qquad (n = 0, \pm 1, \cdots),$$

on the x-axis. Needless to say, the specification of $s(x)$ at the points $x = n + \frac{1}{2}$, which are disregarded in the definition (6), is immaterial, since $s(x)$ occurs in (5) only as an integrable weight-factor, isolated values of which do not influence the integral.

It follows at once from (1), (5) and the continuity (almost everywhere) of (6), that $r(t)$ is continuous.

3. Fourier analysis of $r(t)$. Since the Fourier expansion of the periodic function (6) is

$$s(x) = -\pi^{-1} \sum_{n=1}^{\infty} (-1)^n n^{-1} \sin 2\pi nx = s(x \pm 1) = \cdots, \qquad (|x| < \tfrac{1}{2}), \tag{7}$$

it follows from (5) that[2]

$$r(t) = -\pi^{-1} \sum_{n=1}^{\infty} (-1)^n n^{-1} \int_{-\infty}^{\infty} \phi(x) \sin 2\pi n(x + t)\, dx. \tag{8}$$

Hence, if the sine in (8) is expressed in terms of $2\pi nx$ and $2\pi nt$,

$$\pi r(t) = -\sum_{n=1}^{\infty} (-1)^n n^{-1}(a_n \cos 2\pi nt + b_n \sin 2\pi nt), \tag{9}$$

---

[1] F. Zernike, "Wahrscheinlichkeitsrechnung und mathematische Statistik," *Handbuch der Physik*, Vol. 3 (1928), pp. 475–476.

[2] In view of (1), the term-by-term integration leading from (5) to (8) is justified by the fact that the partial sums of the series (7) are uniformly bounded. Correspondingly, the above deduction of (9) and (10) from (4) is equivalent to an application of Poisson's summation formula. In this regard, cf. A. Wintner, "The sum formulae of Euler-Maclaurin and the inversions of Fourier and Möbius," *Am. Jour. of Math.*, Vol. 69 (1947), pp. 685–708, the end of §1 (p. 687) and its application on p. 697.

where

$$b_n + ia_n = \int_{-\infty}^{\infty} \phi(x) \exp(2\pi inx)\, dx, \qquad (n = 1, 2, \cdots). \tag{10}$$

Let it be assumed that positive and negative errors of observation, when of the same magnitude, are equally probable, i.e., that $\phi(x) = \phi(-x)$. Then (10) shows that a_n becomes 0. Hence, (9) reduces to

$$r(t) = -\sum_{n=1}^{\infty} (-1)^n (c_n/n) \sin 2\pi nt, \tag{11}$$

where

$$c_n = \pi^{-1} \int_{-\infty}^{\infty} \phi(x) \cos 2\pi nx\, dx = 2\pi^{-1} \int_0^{\infty}. \tag{12}$$

Clearly, $r(t)$ is an odd function whenever the density $\phi(x)$ is even.

4. The normal case. Suppose that $\phi(x)$ is the density of a symmetric *normal* (Gaussian) distribution. Then, if σ is the positive constant representing the standard deviation of the errors of observation,

$$\phi(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x^2/\sigma^2) \qquad (0 < \sigma < \infty). \tag{13}$$

It is clear from (5) and (6) that

$$r(t) \to s(t) \text{ if } \sigma \to 0 \text{ in (13)}. \tag{14}$$

Actually, all that (14) says is a triviality, according to which the total error becomes the decimal error when the measurements become infinitely sharp. In this limiting case, that is, if $r(t) = s(t)$, it is seen from (6) that the graph of the periodic function $r = r(t)$ is piecewise linear, and therefore discontinuous.

If $\sigma = 0$ is replaced by $0 < \sigma < \infty$, the jumps of $r(t)$ at $t = n - \frac{1}{2}$ disappear (cf. the end of §3) and, as will be proved below,

(I) *$r(t)$ is an analytic function which is odd, of period* 1, *and positive for* $0 < t < \frac{1}{2}$ (*hence negative for* $-\frac{1}{2} < t < 0$), *and*

(II) *the graph of $r = r(t)$ over the fundamental interval* $0 \leqq t \leqq \frac{1}{2}$ *is a convex arch, no matter what the value of σ in* (13) *may be.*

Since r now depends both on the "true" value, t, of the observations and the "precision", σ, of the measurements, let r be denoted by $r(t; \sigma)$. It will be shown that

(i) Max $r(t; \sigma)$, *where the* Max *refers to t while σ is fixed, is a decreasing function of* σ, where σ varies on the half-line $0 < \sigma < \infty$; and that, on the same half-line,

(ii) *$r(t; \sigma)$ is a decreasing function of σ at every fixed t contained in the fundamental region* $0 < t < \frac{1}{2}$.

All of this seems to be clear for physical reasons. Actually, it is easy to give examples of distribution laws distinct from (13) for which the above assertions become false.

5. The ϑ_3-function. As is well-known,

$$\int_{-\infty}^{\infty} \exp\,(-\tfrac{1}{2}x^2/\sigma^2)\cos ux\,dx = (2\pi\sigma^2)^{\frac{1}{2}}\exp\,(-\tfrac{1}{2}\sigma^2 u^2).$$

Hence, the value of the integral (12) is q^{n^2}, if q is an abbreviation for

$$q = \exp(-2\pi^2\sigma^2). \tag{15}$$

Consequently, if $r(t, q)$ is defined, in terms of the above $r(t; \sigma)$, by placing

$$r(t, q) = r(t; \sigma) \text{ in virtue of (15)}, \tag{16}$$

then (11) shows that[3]

$$r(t, q) = -\pi^{-1}\sum_{n=1}^{\infty}(-1)^n n^{-1} q^{n^2}\sin 2\pi nt \tag{17}$$

It will be noted that the range, $0 < \sigma < \infty$, of the standard deviation is mapped by (15) on the range

$$0 < q < 1, \tag{18}$$

and that σ decreases or increases according as q increases or decreases.

Let partial differentiations with respect to t and q be denoted by primes and subscripts, respectively:

$$f' = \partial f/\partial t, \qquad f_q = \partial f/\partial q. \tag{19}$$

Thus, from (17),

$$r'(t, q) = -2\sum_{n=1}^{\infty}(-1)^n q^{n^2}\cos 2\pi nt \tag{20}$$

and, as easily verified from (17),

$$r_q(t, q) = (-4\pi q)^{-1} r''(t, q). \tag{21}$$

Let $\theta(t, q)$ be defined by

$$\theta(t, q) = 1 + 2\sum_{n=1}^{\infty} q^{n^2}\cos nt \tag{22}$$

(so that $\theta(t, q)$ is, in the main, the elliptic theta-function usually denoted by ϑ_3). It is known that

$$\theta(t, q) > 0 \tag{23}$$

and that[4]

$$\theta'(t, q) < 0 \text{ if } 0 < t < \pi \qquad (\text{hence, } \theta'(t, q) > 0 \text{ if } -\pi < t < 0). \tag{24}$$

The above assertions will be deduced from these facts.

[3] Cf. F. Zernike, loc. cit.

[4] For a simple proof, cf. A. Wintner, "On the shape of the angular case of Cauchy's distribution," *Annals of Math. Stat.*, Vol. 18 (1948), pp. 589–593, §6.

6. Proof of (I)–(II) and (i)–(ii). First, it is seen from (17) and (22) that

$$r'(t, q) = 1 - \theta(2\pi t - \pi, q). \tag{25}$$

Hence,

$$r''(t, q) = -2\pi\theta'(2\pi t - \pi, q). \tag{26}$$

If (26) is compared with (24), it is seen that

$$r''(t, q) < 0 \text{ if } 0 < t < \tfrac{1}{2} \qquad (\text{hence, } r''(t, q) > 0 \text{ if } -\tfrac{1}{2} < t < 0). \tag{27}$$

Consequently, (I) and (II) follow, since, in view of (17),

$$r(\pm\tfrac{1}{2}, q) = 0 = r(0, q). \tag{28}$$

Next, (21) and (27) imply that

$$r_q(t, q) > 0 \text{ for } 0 < t < \tfrac{1}{2}. \tag{29}$$

Hence, (ii) follows from the fact that q is a decreasing function of σ.

As to (i), let $t = t(q)$ denote that (unique) t-value on $0 < t < \frac{1}{2}$ at which $r(t, q)$ assumes its maximum value, say r^q; so that

$$r^q = r(t(q), q), \qquad (0 < t(q) < \tfrac{1}{2}). \tag{30}$$

Clearly, $t = t(q)$ is the only t-value on $0 < t < \frac{1}{2}$ for which

$$r'(t, q) = 0. \tag{31}$$

Since $r'(t, q)$ possesses continuous partial derivatives with respect to t and q, and since (27) implies that its partial derivative with respect to t, namely, $r''(t, q)$, does not vanish at $t = t(q)$, it follows that the solution $t = t(q)$ of the equation (31) possesses a continuous derivative. Hence, the function (30) possesses a continuous derivative with respect to q, namely,

$$\frac{dr^q}{dq} = r'(t(q), q)\frac{dt(q)}{dq} + r_q(t(q), q). \tag{32}$$

But since $t = t(q)$ is a solution of (31), the identity (32) can be reduced to

$$\frac{dr^q}{dq} = r_q(t(q), q), \qquad (0 < t(q) < \tfrac{1}{2}).$$

Consequently, (i) follows from (29), since q is a decreasing function of σ.

WEIGHING DESIGNS AND BALANCED INCOMPLETE BLOCKS

By K. S. Banerjee

Pusa, Bihar, India

1. Introduction. Following a paper by Hotelling [1] on the weighing problem, Kishen [4] and Mood [2] furnished generalized solutions. This note consists of some additional remarks on the weighing problem when the weighing is restricted to be made on one pan.

Hotelling remarked that when the problem was to determine a particular difference or any other linear function of the weights, a different design should be sought to minimize the variance. An account of efficient designs of this kind has also been furnished in this note. The notations used by Hotelling and Mood have been used here.

2. Chemical balance problem. It has been shown by Mood that when $N \equiv 0 \pmod 4$, an optimum design exists if a Hadamard matrix H_N exists, and is obtained by using any p columns of H_N. When $N \equiv i \pmod 4$, $(i = 1, 2, 3)$, very efficient designs are obtained either by adding to or deleting from the rows of H_{4K}, making the resultant number of rows equal to N.

It has further been shown by Mood in connection with this class of designs that arrangements[1] are available which are more efficient than the one obtained by repeating the row of ones. As a matter of fact, if any row other than the row of ones be repeated, this will lead to a design of the same efficiency as in the case of repeated addition of the row of ones; for, the determinant of $X'X$ will remain exactly identical. That this is so, will be clear from the following properties showing the connection of the matrix X with the determinant $|a_{ij}|$:

(i) Any two rows of the matrix X can be interchanged without changing the determinant $|a_{ij}|$.

(ii) Any two columns of the matrix X can be interchanged without changing the determinant $|a_{ij}|$.

(iii) The signs of all the elements in a column of the matrix X may be changed without changing the determinant $|a_{ij}|$.

3. Spring balance problem. Mood has exhaustively discussed the designs when $N > p$. Efficient designs under this class will, however, be available from the arrangements afforded by balanced incomplete block designs discussed in [3]. These designs will be represented by certain of the efficient submatrices of the P_k of Mood.

Usually v and b are used to denote respectively the number of varieties and the number of blocks in the above mentioned designs. Here v will take the place of

[1] This had been independently shown by me before the paper of A. M. Mood was brought to my notice by H. Hotelling.

p, the number of objects to be weighed and b that of N, the number of weighings that can be made. The matrix $X'X$ in this case will take the form

$$
\begin{bmatrix}
r & \lambda & \lambda & \cdots & \lambda \\
\lambda & r & \lambda & \cdots & \lambda \\
\lambda & \lambda & r & \cdots & \lambda \\
\cdot & \cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdot & \cdots & \cdot \\
\lambda & \lambda & \lambda & \cdots & r
\end{bmatrix} \tag{1}
$$

The variance of the estimated weight of each of the p objects for such a design can be easily seen to be

$$\frac{r + \lambda(p - 2)}{(r - \lambda)\{r + \lambda(p - 1)\}}\,\sigma^2 \qquad \text{for zero bias,} \tag{2}$$

where p is the number of objects to be weighed and r and λ have meanings similar to those in connection with balanced incomplete block designs; that is, r is the number of times each object is weighed, and λ is the number of times each pair of objects is weighed together.

Though the *minimum minimorum* of σ^2/N can never be attained by the objects to be weighed under such designs, σ^2/N may however be kept as the standard with which the efficiency of a given design may be calculated. The efficiency of the above design will therefore for zero bias be

$$\frac{(r - \lambda)\{r + \lambda(p - 1)\}}{N\{r + \lambda(p - 2)\}}\,. \tag{3}$$

The identities well known in the theory of balanced incomplete blocks,

$$bk = vr, \qquad \lambda(v - 1) = r(k - 1),$$

may, upon replacing b by N and v by p to accord with the notation of weighing designs, be written

$$r = Nk/p, \qquad \lambda = r(k - 1)/(p - 1).$$

Upon substituting these in (3) we obtain the efficiency factor in the form

$$\frac{k^2(p - k)}{p(pk - 2k + 1)}\,, \tag{4}$$

where k is the number of plots per block or the number of objects that can be weighed at a time.

If instead of adopting repetitions of P_K, only $\binom{p}{K}$ weighings be made in all, the efficiency factor calculated for such a combinatorial design would be

$$\frac{(r - \lambda)\{r + \lambda(v - 1)\}}{b\,\{r + \lambda(v - 2)\}}\,, \qquad \text{for zero bias.}$$

where

$$r = \binom{v-1}{K-1}, \qquad \lambda = \binom{v-2}{K-2}$$

and $b = \binom{v}{K}$. The above expression on simplification reduces to (4).

It will be noticed that the efficiency of such designs depends only upon the total number of objects to be weighed and the number of such objects that can be weighed at a time.

These designs have the advantage that all the weights are estimated with equal precision. If a slightly larger number of weighing than what is afforded by the number of blocks in a balanced incomplete block design has to be made, all the objects may be weighed together and this weighing be repeated as many times as required. This will be equivalent to the repeated addition of the row of ones. The repetition of the row of ones in particular is necessary to make the weights estimable with equal precision, which however, may be demanded at times as a matter of necessity in certain experiments. Otherwise, any other single row or different rows of the matrix X may be repeated, making the number of rows of the matrix X equal to the number of weighings proposed to be made in all.

From the practical point of view also, it will be advantageous to connect the designs for weighing with the already existing balanced incomplete block designs, which have been highly developed in recent years and are being extensively used in agro-biological investigations.

4. Spring balance design for small *p*. Under this class of designs, Mood has found the most efficient design for $p = 7$. It is given by

$$L_7 = \begin{bmatrix} 1&0&1&0&1&0&1 \\ 0&1&1&0&0&1&1 \\ 0&0&0&1&1&1&1 \\ 1&1&0&0&1&1&0 \\ 0&1&1&1&1&0&0 \\ 1&0&1&1&0&1&0 \\ 1&1&0&1&0&0&1 \end{bmatrix}.$$

This L_7 is easily recognized to be the design for $k = 4$, $b = 7$, $v = 7$, $r = 4$, $\lambda = 2$, given by an orthogonal series [3]. It is therefore seen that Hadamard matrices will lead to a new method of constructing balanced incomplete block designs of a certain class. For example H_{16} and H_{20} will lead respectively to the designs for $k = 8$, $b = 15$, $v = 15$, $r = 8$, $\lambda = 4$ (or for $k = 7$, $b = 15$, $v = 15$, $r = 7$, $\lambda = 3$) and for $k = 10$, $b = 19$, $v = 19$, $r = 10$, $\lambda = 5$ (or $k = 9$, $b = 19$, $v = 19$, $r = 9$, $\lambda = 4$). These designs also satisfy the condition of maximum

efficiency, by virtue of the fact that $|L_N|$ will have the value

$$(N+1)^{\frac{1}{2}(N+1)}/2^N,$$

as shown by Mood.

6. Determination of a linear function of the objects. An orthogonalized design which is cent percent efficient to determine individually the weight of p unknown objects is not necessarily the design of maximum efficiency for the estimation of a linear function of the objects. To illustrate this, let there be three objects, the weights O_1, O_2, O_3, of which have to be estimated on a balance corrected for zero bias and let us, for this purpose, concentrate on the design characterized by the matrix given below.

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix}. \tag{5}$$

As has been indicated in the previous papers, the variance of each of the unknown objects comes out to be $\frac{1}{4}\sigma^2$, which is the *minimum minimorum* and as such the above design enjoys the cent percent efficiency, when the question of individual estimation is concerned. But in estimating a linear function of the objects, for instance the total weight, designs more efficient than this are available.

The variance of $l_1O_1 + l_2O_2 + l_3O_3$ is known to be

$$\sum_{i,j=1}^{3} l_i l_j C_{ij} \sigma^2 \tag{6}$$

where C_{ij} denotes the elements of the matrix reciprocal to the matrix $X'X$. As the above design furnishes the estimates of the unknown objects orthogonally, the variance of the estimated total weight of the three objects will be given by $\frac{3}{4}\sigma^2$. If, however, the design given by the matrix

$$x = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \tag{7}$$

be adopted, the variance of the estimate of the total weight may be easily seen to be $(3/7)\sigma^2$, by putting $l_1 = l_2 = l_3 = 1$. $(3/7)\sigma^2$ is evidently less than $\frac{3}{4}\sigma^2$. Therefore with four weighings, the design characterized by (7) is more efficient in estimating the total weight than that characterized by (5). A still more efficient design for getting the total weight is simply to weigh all the objects together four times.

6. Designs with arrangements afforded by balanced incomplete blocks. The necessity for an efficient design to estimate any linear function of the objects

(or to be precise, say to estimate the total weight) will perhaps arise only when the objects cannot all be weighed at a time collectively on a single pan. Here also, an efficient design under the supposition that all the objects cannot be weighed together is afforded by the arrangements in balanced incomplete blocks. In such a design, the diagonal elements in the matrix reciprocal to $X'X$ will be all positive and equal to

$$\frac{r + \lambda(p-2)}{(r-\lambda)\{r + \lambda(p-1)\}}, \tag{8}$$

while the remaining elements in the reciprocal matrix will be negative and equal to

$$\frac{-\lambda}{(r-\lambda)\{r + \lambda(p-1)\}}. \tag{9}$$

Using the generalized form of (6) and admitting of the possibility that any of the arbitrary constants l_i may be negative, the variance of the linear function $\sum_{i=1}^{p} l_i O_i$ may be easily seen to be

$$\left\{\frac{\Sigma l_i^2}{r-\lambda} - \frac{\lambda(\Sigma l_i)^2}{(r-\lambda)\{r+(p-1)\lambda\}}\right\}\sigma^2. \tag{10}$$

If, however, in the above expression, the coefficients l_i are equal to 1, (10) is the variance of the estimated total weight, and reduces to

$$\frac{p}{r + (p-1)\lambda}\sigma^2. \tag{11}$$

When there are N weighings in all, the minimum variance that can be reached is σ^2/N and will be attained, it appears, only when all the objects are weighed together and the weighing is repeated N times. The efficiency of a given design may therefore be calculated with reference to σ^2/N. Remembering that the number of weighings takes the place of the number of blocks and p the place of v, the efficiency of the design will reduce to $(k/p)^2$, where k is the number of plots per block i.e. the number of objects that can be weighed at a time.

If, however, the combinatorial arrangement is adopted weighing all possible combinations of k objects and making $\binom{p}{k}$ weighings in all, the same efficiency as above will be obtained for such a design.

Given k, the above expression of efficiency will therefore be the deciding factor for choice between an arrangement of balanced incomplete block design and all possible combinations of k objects.

7. Design of maximum efficiency. Designs leading to the matrix $X'X$ of the type (1) have certain advantages inasmuch as the variances of the individual objects are equal, as are also the covariances between all possible pairs. The

variance of the estimated total weight in such a design is given by (11). To minimize the variance thus obtained, the expression

$$r + (p - 1)\lambda \tag{12}$$

has to be the maximum for a given value of p. In an arrangement of the balanced incomplete block type or in an arrangement with all possible combinations of k objects being weighed at a time, (12) would reduce to rk and would therefore increase with the increasing value of rk. This shows that the estimation of the total weight will have increased precision if more of the objects are weighed at a time.

If all the objects could be weighed at a time and both the pans be used for the purpose, some of the elements in the matrix X will be -1 instead of 0. This would increase the value of r but would decrease the value of λ. To devise the best possible design therefore, account will have to be taken simultaneously of r and λ.

REFERENCES

[1] HAROLD HOTELLING, "Some improvements in weighing and other experimental techniques", *Annals of Math. Stat.*, Vol. 15 (1944), pp. 297–306.

[2] A. M. MOOD, "On Hotelling's weighing problem", *Annals of Math. Stat.*, Vol. 17 (1946), pp. 432–446.

[3] R. A. FISHER AND F. YATES, *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd, London, 1938, pp. 10–13.

[4] K. KISHEN, "On the design of experiments for weighing and making other types of measurements", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 294–300.

[5] C. R. RAO, "On the most efficient designs in weighing", *Sankhyā*, Vol. 7 (1946), pp. 440

BOUNDS FOR SOME FUNCTIONS USED IN SEQUENTIALLY TESTING THE MEAN OF A POISSON DISTRIBUTION[1]

By Leon H. Herbach

Brooklyn College

1. Introduction. Let $z = \log \dfrac{f(x, \lambda_1)}{f(x, \lambda_0)}$, where $f(x, \lambda_i) = (e^{-\lambda_i} \lambda_i^x)/x!$, $(i = 0, 1)$, is the elementary probability law of a Poisson variate X, under the hypothesis that the mean is equal to λ_i. Without loss of generality we shall assume $\lambda_1 > \lambda_0$.

Let H_0 be the hypothesis that the distribution of X is given by $f(x, \lambda_0)$. Wald [1, pp. 286–287] has devised general upper and lower bounds for the probability of accepting H_0, when λ is the true value of the parameter, and the sequential probability ratio test is used. This probability is called the operating-characteristic function and is designated by $L(\lambda)$. Using these results he has computed the bounds for the binomial and normal distributions [2, pp. 137–142]. We shall do the same thing for the Poisson distribution, since the restrictions [1, p. 284, conditions I to III] under which these general limits are valid can rather easily be shown to apply to the Poisson distribution, if we make the further restriction that $E(z) \neq 0$.

These general results are

$$\frac{1 - B^h}{\delta A^h - B^h} \leq 1 - L(\lambda) \leq \frac{1 - \eta B^h}{A^h - \eta B^h}, \qquad \text{if } h > 0,$$

and

$$\frac{1 - A^h}{\delta B^h - A^h} \leq L(\lambda) \leq \frac{1 - \eta A^h}{B^h - \eta A^h}, \qquad \text{if } h < 0, \tag{1}$$

where α, β are probabilities of committing errors of the first and second kind respectively and

$$\text{A} = (1 - \beta)/\alpha, \qquad \text{B} = \beta/(1 - \alpha)$$

$$\eta = \operatorname*{glb}_{\zeta} \zeta E\left(e^{hz} \mid e^{hz} < \frac{1}{\zeta}\right), \qquad \zeta > 1; \tag{2}$$

$$\delta = \operatorname*{lub}_{\rho} \rho E\left(e^{hz} \mid e^{hz} \geq \frac{1}{\rho}\right), \qquad 0 < \rho < 1;$$

and h is the non-zero root of the expression, $Ee^{zt} = 1$. Hence the only remaining unknowns are η and δ.

---

[1] The author is indebted to Professor A. Wald for suggesting the problem which led to this note and for helpful discussions.

The following bounds to En, the expected number of observations required by the sequential probability ratio test defined by α, β have been derived [1, pp. 143–147]:

$$\frac{L(\lambda)(\log B + \xi') + [1 - L(\lambda)] \log A}{Ez} \begin{matrix} \leq \\ \geq \end{matrix} En$$

$$\begin{matrix} \leq \\ \geq \end{matrix} \frac{L(\lambda) \log B + [1 - L(\lambda)](\log A + \xi)}{Ez},$$

the upper or lower inequality signs holding according as $Ez > 0$ or $Ez < 0$, where

$$\xi' = \operatorname*{Min}_r E(z + r \mid z + r \leq 0), \tag{3}$$

and

$$\xi = \operatorname*{Max}_r E(z - r \mid z - r \geq 0), \qquad (r \geq 0). \tag{4}$$

Using the limits to $L(\lambda)$, we then find ξ and ξ', which determine En.

2. Special terminology. By an *almost-increasing* function we shall mean one that has the following properties: If x is any point of discontinuity, then (a) $x + k$ is also where k is any integer and $x + l$ is a point of continuity if l is not integral, (b) $f(x - \epsilon) < f(x - \epsilon') < f(x)$ for $0 < \epsilon' < \epsilon < 1$, (c) $f(x - 1) < f(x)$, (*d*) $\lim_{\epsilon \to 0} f(x + \epsilon) = f(x +) < f(x)$, (*e*) $f(x - 1 +) < f(x +)$. It is clear that the minimum value for $f(y)$ in any closed interval $[a, b]$ is equal to $\min [f(a), f(a' +)]$ where a' is defined as a if the closed interval contains no discontinuity, and as the leftmost point of discontinuity otherwise. As special cases, if a is a point of discontinuity this minimum is $f(a +)$ and if $x < a < b < x + 1$ the minimum is $f(a)$.

Almost-decreasing functions are defined similarly except that the inequalities go the other way. In this case the maximum in the interval is $\max[f(a), f(a' +)]$ and we have special cases as above.

3. The case $h > 0$. Since $e^z = a^x e^{-c}$, where $a = \lambda_1/\lambda_0$ and $c = (\lambda_1 - \lambda_0)$ the condition $e^{hz} \leq 1/\zeta$ may be expressed as $a^{hx} e^{-ch} \leq 1/\zeta$, whence

$$x \leq c/\log a - \log \zeta/(h \log a) = s - r \text{ (say)}. \tag{5}$$

Since $x \geq 0$, $r \leq s$. Hence $0 < r \leq s$. Also

$$Ee^{zh} = \sum_{x=0}^{\infty} (e^{-c} a^x)^h \frac{e^{-\lambda} \lambda^x}{x!} = \exp(-ch - \lambda + \lambda a^h), \tag{6}$$

and

$$\zeta E(e^{zh} \mid e^{zh} \leq 1/\zeta) = \zeta E[(e^{-c} a^x)^h \mid x \leq s - r]. \tag{7}$$

From (5), $\zeta = a^{rh}$ and (7) becomes

$$(7.1) \qquad a^{rh} \frac{\sum_{x=0}^{[s-r]} \frac{e^{-\lambda}\lambda^x}{x!} e^{-ch} a^{xh}}{\sum_{x=0}^{[s-r]} \frac{e^{-\lambda}\lambda^x}{x!}};$$

where $[s - r]$ is the largest integer $\leq (s - r)$. Our problem is to minimize (7) with respect to ζ. Since r is a strictly increasing function of ζ, this is equivalent to minimizing $a^{rh}C/D = \theta$ (say) with respect to r, where

$$C = \sum_{x=0}^{[s-r]} \frac{\lambda^x a^{xh}}{x!}, \qquad \text{and} \qquad D = \sum_{x=0}^{[s-r]} \frac{\lambda^x}{x!}.$$

It will be shown that (7.1) is an almost-increasing function of r and therefore the minimum occurs at either $r = 0$ or $r = \nu +$, where $\nu = s - [s]$, since the saltuses occur at $r = \nu + k$ for $k = 0, 1, 2, \cdots, [s]$.

Since a^{rh} is an increasing function of r and C/D remains constant as long as $[s - r]$ remains constant, condition (b) is fulfilled.

Conditions (c) to (e) refer to the saltuses only, hence, to show them, we may assume, without loss of generality that r and s are integral. We proceed by induction, using the notation $\theta(w)$ to mean the value of θ, when $r = w$, to show (c).

First we prove the following:

LEMMA A. $\theta(s) > \theta(s - 1)$.

PROOF: Since we assumed $\lambda_1 > \lambda_0$ and $h > 0$, $a^h > 1$. Hence $(1 + \lambda)a^h > 1 + \lambda a^h$, whence, *a fortiori*, $a^{sh} > a^{(s-1)h}(1 + \lambda a^h)/(1 + \lambda)$.

To show that if $\theta(r + 1) > \theta(r)$, then $\theta(r) > \theta(r - 1)$, we shall show that

$$(8) \qquad CD + Dba^{(n+1)h} < CDa^h + Cb$$

implies

$$(9) \qquad CD + Dbqa^{(n+1)h} < CDa^h + Cbqa^h,$$

where $n = s - r$, $b = \lambda^n/n!$, $q = \lambda/(n + 1)$.

Since, as we shall see below,

$$(10) \qquad Dba^{(n+1)h}(q - 1) < Cb(qa^h - 1),$$

or

$$(11) \qquad Da^{(n+1)h}(q - 1) < C(qa^h - 1),$$

addition of (8) and (10) yields the desired result, (9).

It now remains to prove (11) or that

$$(12) \qquad \left[\sum_{x=0}^{n} \frac{\lambda^x}{x!}\right] a^{(n+1)h} (\lambda - n - 1) < \left[\sum_{x=0}^{n} \frac{\lambda^x a^{xh}}{x!}\right] (\lambda a^h - n - 1).$$

Setting (6) equal to 1 we get $\lambda a^h = ch + \lambda$, which when substituted in (12) yields

$$(ch + \lambda)^{n+1}(\lambda - n - 1) \sum_{x=0}^{n} \frac{\lambda^x}{x!} < \lambda^{n+1}(ch + \lambda - n - 1) \sum_{x=0}^{n} \frac{(ch + \lambda)^x}{x!} .$$

Upon letting $p = ch + \lambda$, we have

$$\frac{\lambda - (n + 1)}{\lambda^{n+1}} \sum_{x=0}^{n} \frac{\lambda^x}{x!} < \frac{p - (n + 1)}{p^{n+1}} \sum_{x=0}^{n} \frac{p^x}{x!} = F(p), \qquad \text{say.}$$

Then our problem reduces to showing that $F(y)$ is increasing in $0 < \lambda \le y \le p$ or that the derivative with respect to y, $F'(y)$ is positive.

$$\begin{aligned} F'(y) &= -\sum_{x=0}^{n-1} \frac{(n - x)(y^{x-n-1})}{x!} + (n + 1) \sum_{x=0}^{n-1} \frac{(n - x)(y^{x-n-1})}{(x + 1)!} + (n + 1)^2 y^{-n-2} \\ &> (n + 1)^2 y^{-n-2}, \qquad \text{since } (n + 1) > (x + 1); \\ &> 0 \qquad \text{since } y > 0. \end{aligned}$$

Thus condition (c) is demonstrated. To show (d) we must show that $\theta(r +) < \theta(r)$, which means that

$$a^{rh} \frac{C - ba^{nh}}{D - b} < a^{rh} \frac{C}{D} .$$

But this is true if $C < Da^{nh}$ which is easily verfied. Condition (e) is equivalent to showing that

$$a^{(r-1)h} \frac{C}{D} < a^{rh} \frac{C - ba^{nh}}{D - b} ,$$

which is proved just as (c) was.

Hence,

$$\eta = \min \left\{ e^{-ch} \sum_{x=0}^{[s]} \frac{e^{-\lambda} \lambda^x a^{hx}}{x!} \Big/ \sum_{x=0}^{[s]} \frac{e^{-\lambda} \lambda^x}{x!} , \right. \tag{13}$$
$$\left. a^{\nu h} e^{-ch} \sum_{x=0}^{[s-1]} \frac{e^{-\lambda} \lambda^x a^{hx}}{x!} \Big/ \sum_{x=0}^{[s-1]} \frac{e^{-\lambda} \lambda^x}{x!} \right\}.$$

As special cases we have (i) if s is integral, η is the latter with $\nu = 0$ and (ii) if $s < 1$ (b) is the only applicable condition and we have an ordinary increasing function, hence η is the former.

Similarly, it may be shown that

$$\delta = \max \, [e^{-ch} E(a^{xh} \mid x \ge \{s\}), \quad a^{-\mu h} e^{-ch} E(a^{xh} \mid x \ge \{s + 1\})], \tag{14}$$

where $\{s\}$ is the smallest integer $\ge s$ and $\mu = \{s\} - s$. Here there is only one special case, namely (i). If $h < 0$, δ is the larger of the two expressions on the right side of (13) and η is the smaller of the two corresponding expressions in (14).

4. Since $z = -c + x \log a$, ξ may be written

$$\operatorname*{Max}_t \log aE(x - t \mid x \geq t),$$

where $t = (r + c)/(\log a)$. Hence $s = c/\log a \leq t < \infty$. Therefore if we can show that $E(x - t \mid x \geq t) = \gamma(t)$ (say), is an almost-decreasing function of t we will know that ξ occurs either when $t = s$ or $\{s\}+$ since, as will be seen, the jumps occur at integral t.

To show (c) we make use of the following which is easily proven:

LEMMA B. *Let X, Y, Z each be greater than zero. Then a necessary and sufficient condition that* $\frac{X}{Y} < \frac{X + Y}{Y + Z}$ *is that* $XZ < Y^2$.

Therefore, to show for integral t that

$$\gamma(t) < \gamma(t - 1), \tag{15}$$

or that

$$\frac{\sum_{x=t}^{\infty} (x - t) \frac{\lambda^x}{x!}}{\sum_{x=t}^{\infty} \frac{\lambda^x}{x!}} < \frac{\sum_{x=t}^{\infty} (x - t) \frac{\lambda^x}{x!} + \sum_{x=t}^{\infty} \frac{\lambda^x}{x!}}{\sum_{x=t}^{\infty} \frac{\lambda^x}{x!} + \frac{\lambda^{t-1}}{(t-1)!}},$$

we need only show that, for all integral t,

$$\frac{\lambda^{t-1}}{(t-1)!} \sum_{x=t}^{\infty} \frac{(x - t)\lambda^x}{x!} < \left[\sum_{x=t}^{\infty} \frac{\lambda^x}{x!}\right]^2. \tag{16}$$

Since both sides of (16) are power series in λ where the exponents start with $2t$ we need only show that the coefficient of every term on the left is less than the corresponding term on the right.

In the case of the coefficient of λ^{2j+2t}, $(j \geq 0)$ we have to show that

$$\frac{2j + 1}{(t + 2j + 1)!(t - 1)!} < \frac{2}{(t + 2j)!t!} + \frac{2}{(t + 2j - 1)!(t + 1)!} + \cdots + \frac{1}{(t + j)!(t + j)!},$$

or by multiplying both sides by $(2t + 2j)!$ that

$$(2j + 1)\binom{2t + 2j}{t - 1} < 2\binom{2t + 2j}{t} + 2\binom{2t + 2j}{t + 1} + \cdots + 2\binom{2t + 2j}{t + j - 1} + \binom{2t + 2j}{t + j} = \text{M, say.}$$

Replacing all the binomial coefficients on the right by the smallest one we have

$$(2j + 1)\binom{2t + 2j}{t - 1} < (2j + 1)\binom{2t + 2j}{t} < M,$$

since $\binom{n}{s-1} < \binom{n}{s}$ for $n \geq 2s$. Thus the truth of (16) has been established for even exponents. The odd terms are treated similarly.

Hence, we have shown that $\gamma(t)$ is a strictly decreasing function of t, if t takes on integral values only. We shall now show (b), i.e. that

$$\gamma(t) = \frac{\sum_{x=t}^{\infty} (x - t) \frac{\lambda^x}{x!}}{\sum_{x=t}^{\infty} \frac{\lambda^x}{x!}} < \frac{\sum_{x=\{t-\epsilon\}}^{\infty} (x - t + \epsilon) \frac{\lambda^x}{x!}}{\sum_{x=\{t-\epsilon\}}^{\infty} \frac{\lambda^x}{x!}} = \gamma(t - \epsilon). \tag{17}$$

The denominators are equal and each term of the numerator on the right is greater than the corresponding term on the left, hence (17) is valid.

Conditions (a) and (d) can be shown, by showing in a similar manner, that

$$\gamma(t+) = 1 + \gamma(t + 1) \tag{18}$$

and $\gamma(t) > 1 + \gamma(t + 1)$ for integral t. By using (18) for t and $t - 1$ together with (15) we show $\gamma(t - 1 +) < \gamma(t +)$, which is condition (e). Thus we have shown that

$$\xi = \max \begin{cases} -c + \log a \sum_{x=\{s\}}^{\infty} \frac{x\lambda^x e^{-\lambda}}{x!} \Big/ \sum_{x=\{s\}}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!}, \\ \log a \left[-\{s\} + \sum_{x=\{s+1\}}^{\infty} \frac{x\lambda^x e^{-\lambda}}{x!} \Big/ \sum_{x=\{s+1\}}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} \right]. \end{cases}$$

As in Section 3, ξ' is the lower analogue of ξ, i.e.

$$\xi' = \min \{- c + E(x \mid x \leq [s]), - [s] \log a + E(x \mid x \leq [s - 1])\},$$

and the special cases are as in that section.

REFERENCES

[1] A. Wald, "On cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 283–296.

[2] A. Wald, "Sequential tests of statistical hypotheses," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 117–186.

NOTES

This section is devoted to brief research and expository articles and other short items.

THE DISTRIBUTION OF STUDENT'S t WHEN THE POPULATION MEANS ARE UNEQUAL

By Herbert Robbins

Department of Mathematical Statistics, Universtiy of North Carolina

Let $x_1, \cdots, x_N$ be independent normal variates with the same variance σ^2 and with means $\mu_1, \cdots, \mu_N$ respectively. Set $n = N - 1$ and let

$$\bar{x} = \sum_1^N x_i/N, \qquad s^2 = \sum_1^N (x_i - \bar{x})^2/n, \qquad t = N^{\frac{1}{2}}\bar{x}/s. \tag{1}$$

If all the μ_i are 0 then t has Student's distribution with n degrees of freedom; its frequency function will be denoted here by

$$f_{n,0}(t) = n^{-\frac{1}{2}}\left[B\left(\frac{1}{2}, \frac{n}{2}\right)\right]^{-1} \cdot (1 + t^2/n)^{-\frac{1}{2}(n+1)}. \tag{2}$$

When dealing with situations involving mixtures of populations or in which the mean exhibits a secular trend, it is important to know the distribution of t when the μ_i are arbitrary; in the general case let

$$\begin{aligned} \bar{\mu} &= \sum_1^N \mu_i/N, \qquad & \beta^2 &= \sum_1^N (\mu_i - \bar{\mu})^2/N, \\ \alpha &= N\bar{\mu}^2/2\sigma^2, & \lambda &= N\beta^2/2\sigma^2. \end{aligned} \tag{3}$$

The distribution of t will be shown to depend on the three parameters n, α, λ. If $\lambda = \beta^2 = 0$, so that all the μ_i are equal, then the distribution of t determines the power function of the ordinary t test. We shall here consider the case in which $\alpha = \bar{\mu} = 0$, although the μ_i are different. Denoting the frequency function of t in this case by $f_{n,\lambda}(t)$ we shall show that

$$f_{n,\lambda}(t) = f_{n,0}(t) \cdot \exp\left\{\frac{-\lambda t^2/n}{1 + t^2/n}\right\} \cdot F(-\tfrac{1}{2}, n/2, -\lambda(1 + t^2/n)^{-1}), \tag{4}$$

where F denotes the confluent hypergeometric series, and where, since $\bar{\mu} = 0$,

$$\lambda = \sum_1^N \mu_i^2/2\sigma^2. \tag{5}$$

In fact, the general distribution of t, of which (4) represents the case $\alpha = 0$,

may be derived as follows. Using the standard orthogonal transformation [1, p. 387] let

$$z_i = \sum_{j=1}^{N} c_{ij} x_j, \qquad x_i = \sum_{j=1}^{N} c_{ji} z_j \qquad (i = 1, \cdots, N), \tag{6}$$

where

$$c_{1j} = N^{-\frac{1}{2}} \qquad (j = 1, \cdots, N); \tag{7}$$

then

$$t = n^{\frac{1}{2}} z_1 \Big/ \left(\sum_{2}^{N} z_i^2 \right)^{\frac{1}{2}}. \tag{8}$$

The joint frequency function of the z_i is easily seen to be

$$(2\pi)^{-N/2} \cdot \sigma^{-N} \cdot \exp\left\{ -\sum_{1}^{N} (z_i - a_i)^2 / 2\sigma^2 \right\}, \tag{9}$$

where

$$a_1 = N^{\frac{1}{2}} \bar{\mu}, \qquad \sum_{2}^{N} a_i^2 = N\beta^2. \tag{10}$$

Thus t is the ratio of a non-central normal variate to the square root of an independent non-central chi-square variate. It is known [2, p. 138] that the frequency function of $q^2 = \sum_{2}^{N} z_i^2 / \sigma^2$ is

$$\tfrac{1}{2} e^{-\lambda} \cdot (\tfrac{1}{2} q^2)^{\frac{1}{2}n - 1} \cdot e^{-q^2/2} \cdot \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\lambda q^2)^j}{j!\,\Gamma(\frac{1}{2}n + j)}, \tag{11}$$

where

$$\lambda = \sum_{2}^{N} a_i^2 / 2\sigma^2 = N\beta^2 / 2\sigma^2. \tag{12}$$

The frequency function of $v = z_1/\sigma$ is

$$g(v) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{(\sigma v - a_1)^2}{2\sigma^2} \right\} = \frac{1}{\sqrt{2\pi}} e^{-\alpha} \cdot e^{-(v^2/2)} \cdot \sum_{k=0}^{\infty} \frac{(2\alpha)^{k/2}}{k!} x^k,$$

that of q is, by (11),

$$h(q) = 2^{1-(n/2)} e^{-\lambda} e^{-(q^2/2)} \sum_{j=0}^{\infty} \frac{\lambda^j q^{2j+n-1}}{2^j j!\,\Gamma((n/2) + j)}, \qquad (q > 0),$$

hence that of $u = v/q = n^{-\frac{1}{2}} t$ is

$$\int_0^{\infty} h(q) g(uq) q \, dq,$$

which, after integration, reduces to

$$\pi^{-\frac{1}{2}} e^{-(\lambda+\alpha)} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{\lambda^j (2\alpha^{\frac{1}{2}} u)^k}{j!\,k!} \frac{\Gamma(N/2 + j + k/2)}{\Gamma(n/2 + j)} (1 + u^2)^{-(N+2j+k)/2}. \tag{13}$$

In particular, if $\alpha = \bar{\mu} = 0$ then (13) reduces by means of the relation $F(\alpha, \gamma, x) = e^x F(\gamma - \alpha, \gamma, -x)$ to

$$(14) \quad \left[B\left(\frac{1}{2}, \frac{n}{2}\right)\right]^{-1} \cdot e^{-\lambda u^2/(1+u^2)} \cdot (1 + u^2)^{-\frac{1}{2}N} \cdot F\left(-\frac{1}{2}, \frac{n}{2}, -\lambda(1 + u^2)^{-1}\right),$$

from which it follows that the frequency function of t is given by (4).

Again, let $x_1, \cdots, x_{N_1+N_2}$ be independent normal variates with the same variance σ^2 and with means $\mu_1, \cdots, \mu_{N_1+N_2}$ respectively. Set $n_1 = N_1 - 1$, $n_2 = N_2 - 1$, $n = n_1 + n_2$, and let

$$(15) \quad \begin{gathered} \bar{x}_1 = \sum_1^{N_1} x_i/N_1, \qquad \bar{x}_2 = \sum_{N_1+1}^{N_1+N_2} x_i/N_2 \\ s_1^2 = \sum_1^{N_1} (x_i - \bar{x}_1)^2/n_1, \qquad s_2^2 = \sum_{N_1+1}^{N_1+N_2} (x_i - \bar{x}_2)^2/n_2 \\ s^2 = (n_1 s_1^2 + n_2 s_2^2)/(n_1 + n_2), \qquad t = [N_1 N_2/(N_1 + N_2)]^{\frac{1}{2}}(\bar{x}_1 - \bar{x}_2)/s. \end{gathered}$$

If all the μ_i are equal then t again has Student's distribution with n degrees of freedom. In the general case let

$$(16) \quad \begin{gathered} \bar{\mu}_1 = \sum_1^{N_1} \mu_i/N_1, \qquad \bar{\mu}_2 = \sum_{N_1+1}^{N_1+N_2} \mu_i/N_2, \\ \beta_1^2 = \sum_1^{N_1} (\mu_i - \bar{\mu}_1)^2/N_1, \qquad \beta_2^2 = \sum_{N_1+1}^{N_1+N_2} (\mu_i - \bar{\mu}_2)^2/N_2. \end{gathered}$$

Then we may show as before [1, p. 388] that in this case $u = n^{-\frac{1}{2}}t$ has the frequency function (13), where now

$$(17) \quad \begin{gathered} N = N_1 + N_2 - 1, \qquad \lambda = (N_1\beta_1^2 + N_2\beta_2^2)/2\sigma^2, \\ \alpha = [N_1 N_2/(N_1 + N_2)](\bar{\mu}_1 - \bar{\mu}_2)^2/\sigma^2. \end{gathered}$$

In particular, when $\alpha = \bar{\mu}_1 - \bar{\mu}_2 = 0$, so that $\bar{\mu}_1 = \bar{\mu}_2 = \bar{\mu}$, say, the frequency function $f_{n,\lambda}(t)$ of t is again given by (4), where now

$$(18) \quad \lambda = \sum_1^{N_1+N_2} (\mu_i - \bar{\mu})^2/2\sigma^2.$$

Extensions in this direction to the general linear hypothesis in the analysis of variance will not be treated here

If we set

$$(19) \quad w = (1 + t^2/n)^{-1}$$

where t has the frequency function (4), then w will have the frequency function

$$(20) \quad g_{n,\lambda}(w) = \left[B\left(\frac{1}{2}, \frac{n}{2}\right)\right]^{-1} \cdot e^{-\lambda(1-w)} \cdot w^{\frac{1}{2}n-1} \cdot (1 - w)^{-\frac{1}{2}} \cdot F\left(-\frac{1}{2}, \frac{n}{2}, -\lambda w\right),$$

for $0 < w \leq 1$. Thus for every t,

$$1 - \int_{-t}^{t} f_{n,\lambda}(x)\,dx = \int_0^{(1+t^2/n)^{-1}} g_{n,\lambda}(w)\,dw. \tag{21}$$

It would be interesting to have numerical values of the integral on the left side of (21) for that value of t for which

$$1 - \int_{-t}^{t} f_{n,0}(x)\,dx = 0.01 \text{ or } 0.05 \quad \text{(say)}, \tag{22}$$

but existing tables (e.g. those in [2] and [3]) of the integral of (20) were compiled for a different purpose and do not supply this information. The following remarks throw some light on this subject.

Let us set

$$\begin{aligned} R(t) = f_{n,\lambda}(t)/f_{n,0}(t) &= \exp\left\{\frac{-\lambda t^2/n}{1 + t^2/n}\right\} \cdot F\left(-\frac{1}{2}, \frac{n}{2}, -\lambda(1 + t^2/n)^{-1}\right) \\ &= \{1 - \lambda(t^2/n)/(1 + t^2/n) + o(\lambda)\} \\ &\qquad \cdot \{1 + \lambda/(n + t^2) + o(\lambda)\} \\ &= 1 + \lambda(n + t^2)^{-1}(1 - t^2) + o(\lambda). \end{aligned} \tag{23}$$

Then as $\lambda \to 0$ we have ultimately

$$\begin{aligned} R(t) &> 1 \text{ if } |t| < 1, \\ R(t) &< 1 \text{ if } |t| > 1. \end{aligned} \tag{24}$$

Hence for any $t > 1$ and for sufficiently small λ,

$$1 - \int_{-t}^{t} f_{n,\lambda}(x)\,dx < 1 - \int_{-t}^{t} f_{n,0}(x)\,dx. \tag{25}$$

The exact range of values of t for which $R(t) < 1$ depends of course on n and λ. However we shall show that always

$$R(t) < 1 \text{ if } |t| > 1, \tag{26}$$

so that (25) holds for all n and $\lambda > 0$, provided $t > 1$. The proof is as follows. In terms of w we have

$$R(t) = e^{-\lambda(1-w)} \cdot F(-\tfrac{1}{2}, n/2, -\lambda w) = e^{-\lambda} F((n + 1)/2, n/2, \lambda w). \tag{27}$$

Now

$$\begin{aligned} F((n + 1)/2, n/2, \lambda w) &= 1 \\ &\quad + \sum_{k=1}^{\infty} \frac{(n + 1)(n + 3)\cdots(n + 2k - 1)}{n(n + 2)\cdots(n + 2k - 2)} (\lambda w)^k/k!, \end{aligned} \tag{28}$$

and by induction on k we may show that for all $k = 1, 2, \cdots$,

$$\frac{(n+1)(n+3)\cdots(n+2k-1)}{n(n+2)\cdots(n+2k-2)} \leq 1 + k/n, \tag{29}$$

where the equality holds only for $k = 1$. Hence

$$F((n+1)/2, n/2, \lambda w) < 1 + \sum_{k=1}^{\infty} (1 + k/n)\cdot(\lambda w)^k/k! = e^{\lambda w}(1 + \lambda w/n), \tag{30}$$

$$R(t) < e^{-\lambda(1-w)} \cdot (1 + \lambda w/n) < e^{-\lambda(1-w)} \cdot e^{\lambda w/n} = e^{-\lambda[1-w(1+1/n)]}. \tag{31}$$

Hence $R(t) < 1$ if $w < n/(n+1)$, which is equivalent to (26).

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[2] P. C. Tang, "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat. Res. Memoirs*, Vol. 2 (1938), pp. 127–149.

[3] Emma Lehmer, "Inverse tables of probabilities of errors of the second kind," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 388–398.

A DISTRIBUTION-FREE CONFIDENCE INTERVAL FOR THE MEAN

By Louis Guttman

Cornell University

1. Summary. Consider a random sample of N observations $x_1, x_2, \cdots, x_N$, from a universe of mean μ and variance σ^2. Let m and s^2 be the sample mean and variance respectively:

$$m = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad s^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - m)^2. \tag{1}$$

It is shown that the following conservative confidence interval holds for μ:

$$\text{Prob}\,\{(m-\mu)^2 \leqq s^2/(N-1) + \lambda\sigma^2\sqrt{2/N(N-1)}\} > 1 - \lambda^{-2}, \tag{2}$$

where λ is any positive constant. Inequality (2) also holds if, in the braces, λ is replaced by $\sqrt{\lambda^2 - 1}$, with $\lambda \geqq 1$.

Inequality (2) is much more efficient on the average than Tchebychef's inequality for the mean, namely,

$$\text{Prob}\,\{(m-\mu)^2 \leqq \lambda^2\sigma^2/N\} > 1 - \lambda^{-2}, \tag{3}$$

yet (2) and (3) are both distribution-free, requiring only knowledge about σ^2. At the $1 - \lambda^{-2} = .99$ level of confidence, the expected value of the right member in the braces of (2) is only about 1/6 the corresponding member of (3); at the .999 level of confidence the ratio is about 1/20.

A more general inequality than (2) is developed, also involving only the single parameter σ^2.

2. Derivation. Consider the function

$$u = (m - \mu)^2 - s^2/(N - 1) - c\sigma^2, \tag{4}$$

where c is an arbitrary constant. It is easily verified that $Eu = -c\sigma^2$, and that

$$\underline{E}u^2 = \sigma^4[2/N(N - 1) + c^2]. \tag{5}$$

A basic feature of (5) is that the only population parameter in the right member is σ^2. Contrary to what might have been surmised, the fourth moment of x about μ is not involved, and indeed need not exist.

According to Tchebychef's inequality,

$$\text{Prob}\left\{-\lambda\sqrt{\underline{E}u^2} \leqq u \leqq \lambda\sqrt{\underline{E}u^2}\right\} > 1 - \lambda^{-2}, \tag{6}$$

where λ is an arbitrary positive number. Using (4) and (5), it is possible to write (6) as:

$$\begin{aligned} \text{Prob}\{s^2/(N - 1) + c\sigma^2 - \lambda\sigma^2\sqrt{2/N(N - 1) + c^2} &\leqq (m - \mu)^2 \\ \leqq s^2/(N - 1) + \sigma^2[c + \lambda\sqrt{2/N(N - 1) + c^2}]\} &> 1 - \lambda^{-2}. \end{aligned} \tag{7}$$

In the braces of (7), if the left member is negative, there is no harm in replacing it by zero; if it is positive, then replacing it by zero may only increase the probability of the braces. Regardless of the value of this left member, it is true that

$$\begin{aligned} \text{Prob}\{(m - \mu)^2 \leqq s^2/(N - 1) \\ + \sigma^2[c + \lambda\sqrt{2/N(N - 1) + c^2}]\} > 1 - \lambda^{-2}. \end{aligned} \tag{8}$$

If we set $c = 0$, we have inequality (2). Some improvement over (2) is obtained by determining c to minimize the right member in the braces of (8), yielding as the shortest confidence interval:

$$\text{Prob}\left\{(m - \mu)^2 \leqq s^2/(N - 1) + \sigma^2\sqrt{2(\lambda^2 - 1)/N(N - 1)}\right\} > 1 - \lambda^{-2}. \tag{9}$$

Inequality (9) differs from (2) only by replacing λ in the braces by $\sqrt{\lambda^2 - 1}$.

3. Comparison with Tchebychef's inequality. The expected value of the right member of the braces in (2) is

$$\sigma^2[1/N + \lambda\sqrt{2/N(N - 1)}]. \tag{10}$$

The ratio of (10) to the corresponding value of Tchebychef's inequality (3), namely $\lambda^2\sigma^2/N$, is

$$[1 + \lambda\sqrt{2N/(N - 1)}]/\lambda^2. \tag{11}$$

Since (11) decreases as λ increases, the efficiency of inequality (2) increases compared with that of Tchebychef as the level of confidence $1 - \lambda^{-2}$ increases. The

squared interval of (2) involves only the first power of λ, while that of (3) involves the second power.

4. Approach to normality. If the fourth moment of the universe's distribution exists, then it is well known that the ratio of $\underline{E}(m - \mu)^4$ to σ^4/N^2 must approach 3—the ratio for the normal distribution—as N increases. That is, if $\alpha^2 + 1$ is the ratio, then $\lim_{N\to\infty} \alpha^2 = 2$. It is known[1] that Tchebychef's inequality can be replaced by one involving both α^2 and σ^2, and that

$$\text{Prob}\,\{(m - \mu)^2 \leqq \sigma^2(1 + \lambda\alpha)/N\} > 1 - \lambda^{-2}. \tag{12}$$

If $\alpha^2 = 2$, then the right member in the braces of (12) becomes $\sigma^2(1 + \lambda\sqrt{2})/N$. This is virtually the same as (10), the expected value from (2). In a sense, then, (2) implicitly takes account of the fact that the distribution of sample means approaches that of the normal distribution with respect to the fourth moment. A striking feature, however, is that (2) holds for any $N > 1$ and does not even presume the fourth moment of the universe to exist, whereas to set $\alpha = \sqrt{2}$ in (12) in general requires a large N and finite universe fourth moment.

5. Further possibilities. Confidence interval (2) is derived from but one of a series of general intervals, each of which depends only on σ^2. It may be possible to derive from this series even more efficient intervals, according to the method now to be outlined.

One way of arriving at (2) is to consider all products of the form $(x_i - \mu)(x_j - \mu)$, where $i > j$ and $i, j = 1, 2, \cdots, N$. Let p_2 be the mean of these $N(N - 1)/2$ products. It can easily be seen that $p_2 = u$ in (4) with $c = 0$, so that p_2 is a second degree polynomial in $m - \mu$, the coefficients being sample statistics. A more general quadratic would be $u_2 = p_2 + c_1p_1 + c_0$, where c_1 and c_0 are arbitrary constants and p_1 is the mean of the N values $(x_i - \mu)$ or $p_1 = m - \mu$. It is easily seen that $\underline{E}p_1 = \underline{E}p_2 = \underline{E}p_1p_2 = 0$, and that the only universe parameter involved in $\underline{E}p_1^2$ and $\underline{E}p_2^2$ is σ^2. Hence the only universe parameter upon which u_2^2 depends is also σ^2.

Higher degree polynomials in $m - \mu$ can be defined, possessing the same properties as u_2. Let p_3 be the mean of the $N(N - 1)(N - 2)/3!$ products of the form $(x_i - \mu)(x_j - \mu)(x_k - \mu)$, where $i > j > k$ and $i, j, k = 1, 2, \cdots, N$; etc.; and let $p_N = (x_1 - \mu)(x_2 - \mu) \cdots (x_N - \mu)$. Set $p_0 = 1$, and let

$$u_n = \sum_{a=0}^{n} c_{an}\, p_a \qquad (n = 1, 2, \cdots, N), \tag{13}$$

where the c_{an} are arbitrary constants. It is easily seen that $\underline{E}p_a = 0$ $(a > 0)$, $\underline{E}p_ap_b = 0$ $(a \neq b)$, and that each $\underline{E}p_a^2$ depends on only the parameter σ^2 as far

[1] See, for example, Louis Guttman, "An inequality for kurtosis," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 277–278.

as the universe is concerned. Hence $\underline{E}u_n^2$ depends only on σ^2. Furthermore, by writing $x_i - \mu$ as $(x_i - m) + (m - \mu)$, it is seen that p_a is a polynomial of degree a in $m - \mu$, the coefficients being sample statistics. From (13), then, u_n is a polynomial of degree n in $m - \mu$ with statistics as coefficients.

According to Tchebychef's inequality,

$$\text{Prob}\ \{u_n^2 \leqq \lambda^2 \underline{E}u_n^2\} > 1 - \lambda^{-2}. \tag{14}$$

The interval for u_n^2 in the braces can be expressed in two statements:

$$f_n(m - \mu) = u_n - \lambda\sqrt{\underline{E}u_n^2} \leqq 0, \tag{15}$$

$$g_n(m - \mu) = u_n + \lambda\sqrt{\underline{E}u_n^2} \geqq 0. \tag{16}$$

Both f_n and g_n are polynomials of degree n in $m - \mu$, g_n exceeding f_n always by the additive constant $2\lambda\sqrt{\underline{E}u_n^2}$. Let q_n and Q_n be the smallest and largest real zeros respectively of f_n, and let r_n and R_n be the smallest and largest real zeros respectively of g_n.

For convenience, we can suppose that c_{nn}—the coefficient of $(m - \mu)^n$ in u_n—is positive. If n is even, then f_n is positive for $m - \mu > Q_n$ and for $m - \mu < q_n$. Hence the interval $q_n \leqq m - \mu \leqq Q_n$ contains all the points included in (15) and possibly more. Since the probability of (15) is not less than the probability of (14), we can write the following confidence interval:

$$\text{Prob}\ \{q_n \leqq m - \mu \leqq Q_n\} > 1 - \lambda^{-2} \qquad (n \text{ even}). \tag{17}$$

The problem remains to determine the c_{an} so as to minimize the expected value of $Q_n - q_n$. Inequality (9) provides the minimum for the case $n = 2$. This can be verified by adding the term $c_1 p_1$ to u in (4) and finding that the minimum requires $c_1 = 0$.

If n is odd, we again may set $c_{nn} > 0$. Then $f_n > 0$ for $m - \mu > Q_n$, and $g_n < 0$ for $m - \mu < r_n$. The interval $r_n \leqq m - \mu \leqq Q_n$ thus contains at least all the points found jointly in (15) and (16) and hence forms a conservative confidence interval:

$$\text{Prob}\ \{r_n \leqq m - \mu \leqq Q_n\} > 1 - \lambda^{-2} \qquad (n \text{ odd}). \tag{18}$$

Again, the problem is to determine the c_{an} so as to minimize the expected value of $Q_n - r_n$. Tchebychef's inequality (3) does this for the case $n = 1$.

Although the only *population* parameter involved throughout is σ^2, the *sample* moments up to the nth order are present in (15) and (16). It thus seems plausible that improvement over inequality (9) should be possible for $n > 2$. To obtain such an improvement requires developing a distribution-free theory of the zeros of f_n and g_n beyond the quadratic case.

ON THE COMPOUND AND GENERALIZED POISSON DISTRIBUTIONS

BY E. CANSADO MACEDA

University of Madrid

1. Summary. In this note we deduce several properties of the compound and generalized Poisson distributions; in particular their closure and divisibility properties. An infinite class of functions whose members are both compound and generalized Poisson distributions is exhibited, and several of the distributions of Neyman, Polya, etc. are identified. The present note stems from a paper by Feller [2].

2. The compound Poisson distribution. If $F(x \mid a)$ is a family of distribution functions depending on the parameter a, and $U(a)$ is a distribution function such that it assigns zero probability to any a domain for which $F(x|a)$ is undefined, then

$$G(x) = \int_{-\infty}^{\infty} F(x \mid a)\, dU(a)$$

is a distribution function. In particular if $F(x|a)$ is the Poisson distribution with mean a, and $U(0) = 0$, $G(x)$ is called the *compound Poisson distribution* associated with the distribution function $U(a)$; cf. Feller [2]. Clearly $G(x)$ is a step function over the non-negative integers, the saltus at the point $x = n$ being

$$\pi_n = \int_0^{\infty} e^{-a} \frac{a^n}{n!}\, dU(a), \qquad n = 0, 1, 2, \cdots.$$

It is convenient to introduce the factorial moment generating function (f.m.g.f.) for $G(x)$ as follows

$$\begin{aligned} \omega(z) &= E((1 + z)^x) = \sum_{n=0}^{\infty} \pi_n (1 + z)^n \\ &= \int_0^{\infty} e^{+az}\, dU(a) \\ &= \phi(z) \end{aligned}$$

where $\phi(z)$ is the ordinary moment generating function (m.g.f.) for $U(a)$. This gives a convenient relationship between the moments of $U(a)$ and its associated compound Poisson distribution.

On account of the multiplicative properties of $\omega(z)$ and $\phi(z)$ under the convolution of $G(x)$ and $U(a)$ respectively, it is seen that the compound Poisson distributions form a closed family, and if $G_1(x)$ and $G_2(x)$ are two compound Poisson distributions associated with $U_1(a)$ and $U_2(a)$ respectively then $G_1(x) * G_2(x)$ is associated with $U_1(a) * U_2(a)$. In addition, if $U(a)$ is infinitely divisible (cf. Cramér [1]) then $G(x)$ is also, since it can be factored into the convolution of arbitrarily many compound Poisson distributions.

Choosing in particular $U(a)$ as the Pearson type III distribution, the associated function is the Polya-Eggenberger distribution, and if $U(a)$ is a Poisson distribution the associated function is the Neyman contagious distribution of Type A.

3. The generalized Poisson distribution. If $F(x \mid a)$, defined for non-negative integers $a = 0, 1, 2, \cdots$, is the a-fold convolution of a given distribution $F(x)$ with itself, i.e. $F(x|a) = F(x)^{*a}$, and $U(a)$ is the Poisson distribution with parameter α, then the distribution function

$$G(x) = \int_0^\infty F(x \mid a)\, dU(a)$$

is called the *generalized Poisson distribution* associated with $F(x)$.

If $\Omega(z)$ is the f.m.g.f. of $U(a)$ then for the f.m.g.f. of $G(x)$ we have

$$\omega(z) = \sum_{n=0}^{\infty} (\Omega(z))^n e^{-\alpha} \frac{\alpha^n}{n!}$$
$$= e^{\alpha(\Omega(z)-1)}.$$

It follows that $\omega(z)$ can be written as $\prod_{\nu=1}^{\infty} \omega_\nu(z)$ where $\omega_\nu(z)$ is a generalized Poisson distribution, and thus $\omega(z)$ belongs to the infinitely divisible family. Moreover, if $G_1(x)$ and $G_2(x)$ are two generalized Poisson distributions associated with $U_1(a)$ and $U_2(a)$ with parameters α_1 and α_2 respectively, then $G(x) = G_1(x)_* G_2(x)$ has for f.m.g.f.

$$\omega_1(z)\omega_2(z) = \exp\left\{(\alpha_1 + \alpha_2)\left(\frac{\alpha_1\Omega_1(z) + \alpha_2\Omega_2(z)}{\alpha_1 + \alpha_2} - 1\right)\right\},$$

and $G(x)$ is again a generalized Poisson distribution function associated with the distribution

$$U(a) = \frac{\alpha_1 U_1(a) + \alpha_2 U_2(a)}{\alpha_1 + \alpha_2}$$

and with the parameter $\alpha_1 + \alpha_2$. Thus the generalized Poisson distributions form a closed family. The analytic nature of the generalized Poisson distributions have been studied by Hartman and Wintner [3]. As noted by Feller [2] the various Neyman contagious distributions are generalized Poisson distributions.

4. Further remarks. From the above observations it is clear that a necessary and sufficient condition for a distribution to be a compound Poisson distribution is that its f.m.g.f. be of the form

$$\omega_1(z) = \phi(z) \tag{1}$$

where $\phi(z)$ is the ordinary m.g.f. of a non-negative random variable. Likewise a necessary and sufficient condition for $\omega(z)$ to be the f.m.g.f. of a generalized Poisson distribution is that it be of the form

$$\omega_2(z) = e^{\alpha(\Omega(z)-1)}, \qquad \alpha > 0, \tag{2}$$

where $\Omega(z)$ is the f.m.g.f. of an arbitrary distribution function $F(x)$. If we choose $\phi(z) = e^{\alpha(e^{cz}-1)}$ and $\Omega(z) = e^{cz}$, then $\omega_1(z) = \omega_2(z)$, and the distribution whose f.m.g.f. is $\omega_1(z)$ (the Neyman contagious distribution of Type A) is simultaneously a compound and a generalized Poisson distribution (cf. Feller [2]). We now show that there is an infinite class of distributions with this property.

First note that if $\phi(z)$ is the m.g.f. of an arbitrary distribution, then exp $\{\alpha(\phi(z) - 1)\}$ is also the m.g.f. of a d.f., and in fact is the m.g.f. of the generalized Poisson distribution associated with the distribution whose m.g.f. is $\phi(z)$. Now let $\phi(z)$ be the m.g.f. of an arbitrary non-negative random variable, and define

$$\omega(z) = \exp\{\alpha(\phi(z) - 1)\} \qquad \alpha > 0. \tag{3}$$

Then $\omega(z)$ is simultaneously of the forms (1) and (2), since $\phi(z)$ is, by (1), also the f.m.g.f. of a distribution function, i.e. the compound Poisson distribution associated with the distribution whose m.g.f. is $\phi(z)$. However, not every distribution which is both a compound and a generalized Poisson distribution can be generated in this manner. For example, the Polya-Eggenberger distribution is easily shown to be both a generalized and a compound Poisson distribution, yet its f.m.g.f.

$$\omega(z) = (1 - dz)^{-h/d}, \qquad d > 0, h > 0,$$

manifestly is not of the form (3), since this would imply $\phi(iz) = 1 - \frac{h}{\alpha d}\log(1 - diz)$ is a characteristic function. But $|\phi(iz)|$ is unbounded as $z \to \pm\infty$ and thus is not the characteristic function of a distribution.

REFERENCES

[1] H. Cramér, "Problems in probability theory," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 165–193.

[2] W. Feller, "On a general class of contagious distributions," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 389–400.

[3] P. Hartman and A. Wintner, "On the infinitesimal generators of integral convolutions," *Am. Jour. of Math.*, Vol. 64 (1942), pp. 272–279.

ON CONFIDENCE LIMITS FOR QUANTILES

By Gottfried E. Noether

Columbia University

In finding confidence limits for quantiles it is usual to determine two order statistics Z_i and Z_j which with a given probability contain the unknown quantile

between them. The values of i and j corresponding to a given confidence coefficient can be determined with the help of the distribution laws of order statistics as is shown, e.g., in Wilks [1]. The purpose of this note is to determine i and j with the help of a confidence band for the unknown cumulative distribution function.

In what follows we shall always denote the cumulative distribution function (cdf) by $F(x)$, i.e., $F(x) = P\{X \leq x\}$. Then the quantile q_p is determined by

$$F(q_p - 0) \leq p \leq F(q_p) \tag{1}$$

which reduces to

$$F(q_p) = p \tag{1'}$$

if $F(x)$ is continuous. Given a sample of size n we can construct the sample cdf $F_n(x)$ defined by $F_n(x) = 1/n$ (number of observations $\leq x$). Confidence coefficients will always be denoted by $1 - \alpha$.

Assume that we can construct two step functions $L(x)$ and $U(x)$ parallel to $F_n(x)$ such that for any fixed value x

$$P\{L(x) < F(x) < U(x)\} = 1 - \alpha. \tag{2}$$

We do not require that the confidence band determined by $L(x)$ and $U(x)$ cover the graph of the unknown cdf $F(x)$ with probability $1 - \alpha$, but only that for any arbitrarily chosen value x (2) is true.

Let

$$L(x) = \eta_k, \qquad U(x) = \theta_k$$

for $z_k \leq x < z_{k+1}$, $k = 0, 1, \cdots, n$ where z_k is the value taken by the order statistic Z_k and $z_0 = -\infty$, $z_{n+1} = +\infty$. Then if $F(x)$ is continuous it follows from (2) that a confidence interval with confidence coefficient $1 - \alpha$ for q_p is given by

$$Z_i \leq q_p < Z_j \tag{3}$$

where i and j are determined by

$$\theta_{i-1} \leq p, \qquad \theta_i > p \tag{4}$$

$$\eta_{j-1} < p, \qquad \eta_j \geq p. \tag{5}$$

It will be noted that (3) represents a half-open interval. However as long as we only admit continuous cdf's the confidence coefficient is not changed if we use

$$Z_i < q_p < Z_j \tag{3'}$$

or

$$Z_i \leq q_p \leq Z_j \tag{3''}$$

instead. This is no longer true if we also admit discontinuous cdf's. Then the confidence coefficient connected with (3′) is $\leq 1 - \alpha$, while that connected with

(3″) is $\geq 1 - \alpha$, as follows immediately from consideration of the possible outcomes when (1) is true. This is the same result as that obtained by Scheffé and Tukey [2].

We shall now indicate how η_k and θ_k can be obtained and find their values in a particular case. For any arbitrary value x we can consider $F_n(x)$ as the sample estimate of the unknown parameter $p = F(x)$ of a binomial distribution. Clopper and Pearson [3] have discussed how confidence intervals for the unknown parameter of a binomial variate can be found. Thus we can determine η_k and θ_k correspondingly, but as is well known (2) cannot be achieved with probability exactly equal to $1 - \alpha$. We shall have to be satisfied with probability $\geq 1 - \alpha$. Consequently the same will hold true for the confidence coefficient connected with the confidence interval for q_p.

In many cases central confidence intervals seem to be more desirable, at least intuitively, than others. Our method produces such central confidence intervals for the unknown quantile if we use central confidence intervals in the construction of the confidence band. In that case η_k and θ_k are determined by

$$\begin{aligned} &(6) \qquad \frac{\alpha}{2} = I_{\eta_k}(k, n - k + 1) \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad k = 0, 1, \cdots, n \\ &(7) \qquad \frac{\alpha}{2} = I_{1-\theta_k}(n - k, k + 1), \end{aligned}$$

except that $\eta_0 = 0$, $\theta_n = 1$ by definition, where

$$I_x(p, q) = \int_0^x t^{p-1}(1 - t)^{q-1}\, dt \Big/ \int_0^1 t^{p-1}(1 - t)^{q-1}\, dt$$

is the incomplete beta function. Scheffé [4] has pointed out how the tables of percentage points of the incomplete beta function by C. M. Thompson, etc. [5] can be used to find η_k and θ_k.

We shall show now that in the case of the median M the solution based on (3)–(7) leads to the same confidence interval as that suggested originally by W. R. Thompson [6]. Thompson found that for $k < n + \frac{1}{2}$

$$(8) \qquad P\{Z_k < M < Z_{n-k+1}\} = 1 - 2I_{\frac{1}{2}}(n - k + 1, k)$$

provided the unknown distribution had a continuous cdf. (8) can be used to maximize k under the condition that the righthand side is $\geq 1 - \alpha$.

We shall first show that our method leads to the same kind of a confidence interval, i.e., one with $i = l$, $j = n - l + 1$. This follows immediately from the fact that by (6) and (7)

$$(9) \qquad 1 - \theta_l = \eta_{n-l}.$$

For let

$$(10) \qquad \theta_{l-1} \leq \tfrac{1}{2} \text{ and } \theta_l > \tfrac{1}{2},$$

then by (9) $\eta_{n-l} < \frac{1}{2}$ and $\eta_{n-l+1} \geq \frac{1}{2}$.

It remains to be shown that k as determined by (8) equals l. This will be so if we can show that

$$I_{\frac{1}{2}}(n - l + 1, l) \leq \frac{\alpha}{2} < I_{\frac{1}{2}}(n - l, l + 1). \tag{11}$$

Remembering that $I_x(p, q)$ is a monotonically increasing function of x we get with the help of (7) and (10)

$$\frac{\alpha}{2} = I_{1-\theta_{l-1}}(n - l + 1, l) \geq I_{\frac{1}{2}}(n - l + 1, l)$$

and

$$\frac{\alpha}{2} = I_{1-\theta_l}(n - l, l + 1) < I_{\frac{1}{2}}(n - l, l + 1)$$

which proves (11).

In conclusion it may be worth while pointing out that the formula

$$P\{Z_i < q_p < Z_j\} = I_p(i, n - i + 1) - I_p(j, n - j + 1)$$

given, e.g, in Wilks [1] for the continuous case can be obtained by a slight modification of (6).

REFERENCES

[1] S. S. Wilks, "Order statistics," *Am. Math. Soc. Bull.*, Vol. 54 (1948), pp. 6-50.
[2] H. Scheffé and J. W. Tukey, "Non-parametric estimation. I. Validation of order statistics," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 187-192.
[3] C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, Vol. 26 (1934), pp. 404-413.
[4] H. Scheffé, "Note on the use of the tables of percentage points of the incomplete beta function to calculate small sample confidence intervals for binomial p," *Biometrika*, Vol. 33 (1944), p. 181.
[5] C. M. Thompson, E. S. Pearson, L. J. Comrie, and H. O. Hartley, "Tables of percentage points of the incomplete beta function," *Biometrika*, Vol. 32 (1941), pp. 151-181.
[6] W. R. Thompson, "On confidence ranges for the median and other expectation distributions for populations of unknown distribution form," *Annals of Math. Stat.*. Vol. 7 (1936), pp. 122-128.

A LOWER BOUND FOR THE EXPECTED TRAVEL AMONG m RANDOM POINTS

By Eli S. Marks

Bureau of the Census

In connection with cost determinations in sampling problems, it is frequently necessary to determine the amount of travel among m random sample points in an area. A lower bound for the expected value of this distance is found to be:

$$\sqrt{\frac{A}{2}}\,\frac{m-1}{\sqrt{m}},$$

where A is the measure of the area from which the m random points are drawn.[1]

If in a finite area S we locate m points at random (see Figure 1), we can trace a continuous path among the m points by starting at some point and connecting the points by line segments. The points can be connected in any order so that the path touches each point only once (unless it intersects itself at one of the random points). We are interested in a lower bound for the expected value of the length of the shortest of the $m!$ possible paths.

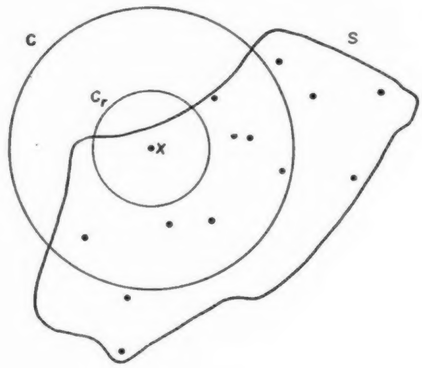

FIG. 1. m RANDOM POINTS IN S.

We have above an area S in which m random points have been selected (with $m = 14$).

The shortest path among the m points consists of $m - 1$ "links" (line segments) between two points. Each link can be assigned to one of its end points, leaving some pre-designated point (e.g., the m-th point selected) with no link assigned. The link assigned to the i-th random point ($x_{(i)}$) must be no less than $r_{(i)}$ the distance from $x_{(i)}$ to the nearest of the other $(m - 1)$ points. If we denote the length of the shortest path by L:

$$L \geq \sum_{i=1}^{m-1} r_{(i)},$$

$$E(L) \geq \sum_{i=1}^{m-1} E(r_{(i)}).$$

Let $E_x(r_{(i)})$ be the expected value of $r_{(i)}$ conditional upon $x_{(i)}$ falling at the point x in S and let $F(r \mid x)$ be the conditional distribution function of $r_{(i)}$ for $x_{(i)} = x$. Thus $F(r \mid x)$ is the conditional probability of $r_{(i)} \leq r$ or the probability of

[1] The lower bound obtained is similar in form to the expression for distance traveled among a set of random points used by Mahalanobis [2] and Jessen [1].

one or more of the $(m - 1)$ random points other than $x_{(i)}$ falling inside a circle, C_r, with radius r and center at x (see Fig. 1). Then, we have:

$$E_x(r_{(i)}) = \int_0^{+\infty} r\,dF(r|x),$$

$$F(r \mid x) = 1 - \left\{\frac{M(S) - M(SC_r)}{M(S)}\right\}^{m-1},$$

where $M(S)$ and $M(SC_r)$ are the measures of S and SC_r, so that $\dfrac{M(SC_r)}{M(S)}$ is the probability of a random point in S falling into C_r.

Let $A = M(S)$ and construct a circle C with center at x and radius $\rho = \sqrt{\dfrac{A}{\pi}}$. Then $M(C) = A = M(S)$. Let d be the distance from x to the nearest of $(m - 1)$ points selected at random from C and let $G(r)$ be the distribution function of d. Then we have:

$$E(d) = \int_0^{+\infty} r\,dG(r),$$

$$G(r) = 1 - \left\{\frac{M(C) - M(CC_r)}{M(C)}\right\}^{m-1}.$$

For $r \leq \rho$,

$$M(C_rC) = M(C_r) \geq M(SC_r).$$

For $r > \rho$,

$$M(C_rC) = M(C) = M(S) \geq M(SC_r).$$

Thus, since $M(C_rC) \geq M(SC_r)$, we have for all x in S:

$$G(r) \geq F(r \mid x),$$

and thus,

$$E(d) \leq E_x(r_{(i)}).$$

Since $E(d) \leq E_x(r_{(i)})$ for all x in S:

$$E(d) \leq E(r_{(i)}),$$

$$(m - 1)E(d) \leq \sum_{i=1}^{m-1} E(r_{(i)}) \leq E(L).$$

It only remains to evaluate $E(d)$, the expected distance from the center of a circle to the nearest of $(m - 1)$ random points. This can be done very easily by substituting in the expression for $G(r)$:

$$A = M(C),$$

$$\pi r^2 = M(C_rC), \qquad \text{when } r \leq \rho = \sqrt{\frac{A}{\pi}},$$

to give:

$$G(r) = 1 - \left\{\frac{A - \pi r^2}{A}\right\}^{m-1},$$

$$G'(r) = \frac{2\pi r}{A}(m - 1)\left\{\frac{A - \pi r^2}{A}\right\}^{m-2},$$

$$E(d) = \int_0^\rho rG'(r)\,dr = \frac{1}{2}\sqrt{\frac{A}{\pi}}\,[B(m, \tfrac{1}{2})],$$

where $B(m, \frac{1}{2})$ is the complete Beta function.
Since $\sqrt{m}\,[B(m, \frac{1}{2})] \geq \sqrt{\pi}$:

$$E(d) \geq \frac{1}{2}\sqrt{\frac{A}{m}}.$$

Thus, we have:

$$E(L) \geq \tfrac{1}{2}\sqrt{A}\,\frac{m - 1}{\sqrt{m}}.$$

It is obvious that the development is general and applies to m random points in any bounded two-dimensional Borel set. However, the lower bound obtained will, in general, be useful only when S is a connected region.

REFERENCES

[1] RAYMOND J. JESSEN, "Statistical investigation of a sample survey for obtaining farm facts," *Iowa State College Research Bulletin* 304 (1942).
[2] P. C. MAHALANOBIS, "A sample survey of the acreage under jute in Bengal," *Sankhyā*, Vol. 4 (1940), pp. 511–530.

A MATRIX ARISING IN CORRELATION THEORY[1]

BY H. M. BACON

Stanford University

1. Introduction. In the study of time series, it is frequently desirable to consider correlations between observations made in different years. Let x_{i1}, x_{i2}, $\cdots$, x_{im} be m values of the variable x_i, expressed as deviations from their arithmetic mean, where x_i is a variable observed in the ith year ($i = 1, 2, \cdots, n$).

[1] A linear correlogram is considered by Cochran in his paper, "Relative accuracy of systematic and stratified random samples for a certain class of populations," (*Annals of Math. Stat.*, Vol. 17 (1946), pp. 164–177) in which $\rho_\mu = 1 - \frac{\mu}{L}$. Setting $\mu = |i - j|$ and $L = 1/p$, we have the case considered above.

Let σ_i be the standard deviation of x_i. If we denote by $r_{ij} = r_{ji}$ the correlation of x_i with x_j, and if we assume the x_i to be normally distributed, then

$$z = \frac{1}{(2\pi)^{n/2}\sigma_1\sigma_2 \cdots \sigma_n\sqrt{R}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{R_{ij}x_i x_j}{R\sigma_i\sigma_j}\right\}$$

is the frequency function giving the distribution. Here R is the determinant $|r_{ij}|$ of the correlation coefficients, and R_{ij} is the cofactor of the element r_{ij} in this determinant.

We may make various assumptions regarding the behavior of the correlation coefficients over the n years. One such assumption of some interest is that the correlation coefficients diminish in such a way that

$$r_{ij} = r_{ji} = 1 - |i - j|p$$

where p is a fixed positive number not greater than $2/(n - 1)$. Under these circumstances, we can evaluate R and R_{ij} in terms of n and p.

2. Evaluation of R. We may let $R(p)$ represent the determinant R of order n whose element in the ith row and jth column is $r_{ij} = r_{ji} = r_{n-i,n-j} = r_{n-j,n-i} = 1 - |i - j|p$ where, for the purpose of evaluation, p is any real number. Since each two-rowed minor of $R(p)$ is divisible by p, $R(p)$ is divisible by p^{n-1}. Furthermore, since $R(p)$ is a polynomial in p of degree at most n, we have

$$R(p) = Ap^n + Bp^{n-1} = p^{n-1}(Ap + B).$$

If we set $p = 1$ and $p = -1$, we find $A + B = R(1)$ and $R(-1) = (-1)^{n-1}(-A + B)$ so that $-A + B = (-1)^{n-1}R(-1)$. By elementary methods we find that $R(1) = 2^{n-2}(3 - n)$ and $R(-1) = (-1)^{n-1}2^{n-2}(n + 1)$. Hence

$$A + B = 2^{n-2}(3 - n)$$

and

$$-A + B = 2^{n-2}(n + 1).$$

Solving for A and B we find that

$$R = R(p) = 2^{n-2}p^{n-1}[2 - (n - 1)p].$$

3. Evaluation of R_{ij}. Similar methods yield the following values for the cofactors R_{ij} of the elements of R:

$$\begin{aligned} R_{11} = R_{nn} &= 2^{n-3}p^{n-2}[2 - (n - 2)p], \\ R_{22} = R_{33} &= \cdots = R_{n-1,n-1} = 2^{n-2}p^{n-2}[2 - (n - 1)p], \\ R_{1n} = R_{n1} &= 2^{n-3}p^{n-1}, \\ R_{i,i+1} &= -2^{n-3}p^{n-2}[2 - (n - 1)p], \end{aligned}$$

otherwise,

$$R_{ij} = 0.$$

4. The frequency function. The quadratic form appearing in the exponent in the expression for the frequency function can now be written as

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{R_{ij}x_i x_j}{R\sigma_i\sigma_j} = \frac{2-(n-2)p}{2p[2-(n-1)p]}\left(\frac{x_1^2}{\sigma_1^2}+\frac{x_n^2}{\sigma_n^2}\right)$$
$$+\frac{1}{p}\left(\frac{x_2^2}{\sigma_2^2}+\frac{x_3^2}{\sigma_3^2}+\cdots+\frac{x_{n-1}^2}{\sigma_{n-1}^2}\right)$$
$$+\frac{1}{2[2-(n-1)p]}\left(\frac{x_1x_n}{\sigma_1\sigma_n}+\frac{x_nx_1}{\sigma_n\sigma_1}\right)$$
$$-\frac{1}{2p}\left(\frac{x_1x_2}{\sigma_1\sigma_2}+\frac{x_2x_1}{\sigma_2\sigma_1}+\frac{x_2x_3}{\sigma_2\sigma_3}+\frac{x_3x_2}{\sigma_3\sigma_2}+\cdots+\frac{x_nx_{n-1}}{\sigma_n\sigma_{n-1}}\right)$$
$$=\frac{1}{p}\left[\frac{2-(n-2)p}{2[2-(n-1)p]}\left(\frac{x_1^2}{\sigma_1^2}+\frac{x_n^2}{\sigma_n^2}\right)+\sum_{i=2}^{n-1}\frac{x_i^2}{\sigma_i^2}-\sum_{i=1}^{n-1}\frac{x_i x_{i+1}}{\sigma_i\sigma_{i+1}}\right]$$
$$+\frac{1}{2-(n-1)p}\left(\frac{x_1x_n}{\sigma_1\sigma_n}\right).$$

5. Maximum likelihood. The expression z is the likelihood of getting a particular set of values of the variables $x_1, x_2, \cdots, x_n$. It is often important to regard the r_{ij} and the σ_i as parameters and to determine them so that the likelihood will be a maximum. If we assume $\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$, then

$$z = \frac{1}{(2\pi)^{n/2}\sigma^n\sqrt{R}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{R_{ij}x_ix_j}{R\sigma^2}\right\}.$$

The question, in our case, now becomes, What values of p and σ will make z a maximum for given x_i? Necessary conditions are that $\frac{\partial z}{\partial p} = 0$ and $\frac{\partial z}{\partial \sigma} = 0$. Since R_{ij} and R are given in terms of p, the process of differentiation can be carried out (first take the logarithm of z), and values of p and σ necessary for a maximum determined. It is, of course, possible that z has no maximum, and the sufficiency of these values must be tested. The computations for the general case are laborious, though straightforward. Furthermore, because of the complicated nature of the coefficients in the equation to be solved for p, the general solution is not readily obtainable. This equation is, however, of third degree, and it can be solved in any particular case.

TABLE OF NORMAL PROBABILITIES FOR INTERVALS OF VARIOUS LENGTHS AND LOCATIONS

By W. J. Dixon

University of Oregon

1. Introduction. The probability associated with a particular finite range of values is often desired. The usual tables of normal areas gives values for $\int_0^x$ or

as in the table by Salvosa [1], $\int_{-\infty}^{x}$. The WPA table [2] gives $\int_{-x}^{x}$. The author has deposited with Brown University a table of $\int_{x-\frac{1}{2}l}^{x+\frac{1}{2}l}$ for values of $x[0(.1)\ 5.0]$ and values of $l[0(.1)\ 10.0]$. The values in the table may be interpreted as the probability that an observation from a normal population with unit variance will fall in an interval of length l whose midpoint is a distance x from the mean. These values can be obtained by a simple computation from the existing tables. Since values were being used frequently, the present table was constructed. Microfilm or photostat copies may be obtained upon request to the Brown University Library.

2. Computation. The values were obtained by finding the difference between the integrals $\int_{-\infty}^{x-\frac{1}{2}l}$ and $\int_{-\infty}^{x+\frac{1}{2}l}$ as given to six decimal places in Salvosa's table. Being differences, the values are subject to an error of 1 unit in the sixth place. For values of $x + \frac{1}{2}l$ greater than 5, the values can be obtained by computing $1 - \int_{-\infty}^{x-\frac{1}{2}l}$. The search for errors was aided by computing column sums; i.e.

$$\sum_{i=1}^{50} \int_{x_i-\frac{1}{2}l}^{x_i+\frac{1}{2}l} + \frac{1}{2}\int_{-\frac{1}{2}l}^{\frac{1}{2}l} = .5\ n, \tag{1}$$

where i represents the row number and n represents the column number. For example, $n = 17$ corresponds to column for $l = 1.7$. The approximation becomes poorer as n increases but the sums were still useful for checking purposes.

3. Example. The table has been used in studies of the expected proportion of a line covered by intervals dropped on it according to some normal probability function. Let $P_n(x)$ be the probability that the point x is covered at least once when n intervals are dropped on the x-axis. H. E. Robbins [3] gives the expression:

$$E(F) = \frac{1}{L}\int_0^L P_n(x)\ dx, \tag{2}$$

for the expected proportion of a line of length L covered at least once by these intervals.

Let $f(x)\ dx$ be the probability that an interval falls with its center in dx and l be the length of the interval. The probability that a point x will be covered by one interval dropped on the x-axis is:

$$g(x) = \int_{x-\frac{1}{2}l}^{x+\frac{1}{2}l} f(t)\ dt. \tag{3}$$

When n intervals are dropped, the probability that x is covered at least once is:

$$P_n\ (x) = 1 - (1 - g(x))^n, \tag{4}$$

and

$$E(F) = 1 - \frac{1}{L}\int_0^L (1 - g(x))^n \, dx. \tag{5}$$

When k groups of n_i intervals are dropped according to, say normal distributions with different means,

$$P_n(x) = 1 - \prod_{i=1}^{k} (1 - g_i(x))^{n_i}. \tag{6}$$

Where

$$g_i(x) = \int_{x-\frac{1}{2}l}^{x+\frac{1}{2}l} f_i(t) \, dt \tag{7}$$

and we obtain

$$E(F) = 1 - \frac{1}{L}\int_0^L \prod_{i=1}^{k} (1 - g_i(x))^{n_i} \, dx. \tag{8}$$

The values $g(x)$ are those given in the table and are useful in evaluating the integrals in (5) and (8) by numerical methods.

REFERENCES

[1] Luis R. Salvosa, "Tables of Pearson's Type III functions," *Annals of Math. Stat.*, Vol. 1 (1930), p. 191.

[2] National Bureau of Standards, *Tables of Probability Functions*, Vol. 2 (1942).

[3] H. E. Robbins, "On the measure of a random set," *Annals of Math. Stat.*, Vol. 15, (1944).

CORRECTION TO "A NOTE ON THE FUNDAMENTAL IDENTITY OF SEQUENTIAL ANALYSIS"

By G. E. Albert

University of Tennessee

In the paper cited in the title (*Annals of Math. Stat.*, Vol. 18 (1947), pp. 593–596), the proof of Lemma 3 is incorrect. The following correct proof is due to Mr. C. R. Blyth of the Institute of Statistics, University of North Carolina.

It is easy to establish the equation

$$P(n = N|F)[\varphi(t_0)]^{-N} = P(n = N|G)E_{n=N}[\exp(-t_0Z_N)|G],$$

where $E_{n=N}(u|G)$ denotes the conditional expectation of u under the condition that $n = N$ for any fixed integer N. By Wald [2], equations (2.4) and (2.6), there exists a finite constant C independent of N which dominates the expected values $E_{n=N}[\exp(- t_0Z_N)|G]$ for every N. Thus

$$P(n = N|F)[\varphi(t_0)]^{-N} \leqq C\cdot P(n = N|G). \tag{A}$$

By Stein's theorem [3], there is a positive number t_1 such that $E(\exp nt_1|G)$ is finite. But by (A),

$$E\{\exp n[t_1 - \log \varphi(t_0)]\} \leqq C \cdot E(\exp nt_1|G),$$

and Lemma 3 is proved.

CORRECTION TO "ON THE CHARLIER TYPE B SERIES"

By S. Kullback

George Washington University

In the paper cited in the title (*Annals of Math. Stat.*, Vol. 18 (1947), p. 575), the phrase "so that . . . $R_1 > 1$" on lines 5 and 6 should be deleted. I am grateful to Prof. Ralph P. Boas, Jr. for calling this to my attention.

ABSTRACTS OF PAPERS

Presented June 22–24, 1948 at the Berkeley Meeting of the Institute

1. Estimation of Parameters for Truncated Multinormal Distributions. Z. W. Birnbaum, E. Paulson and F. C. Andrews, University of Washington.

Let $X_{(N)} = (X_1, \cdots X_p, X_{p+1}, \cdots, X_N)$ be an N-dimensional random variable with a non-singular normal distribution, and let the expectations, variances and covariances of $X_{p+1}, \cdots X_N$ be known. A large sample of $X_{(N)}$ is available, obtained under some side-condition on $(X_{p+1}, \cdots, X_N)$; this side-condition may be a truncation of any kind or, more generally, a selection; i.e. imposing on $X_{p+1}, \cdots, X_N$ a probability-distribution different from the original marginal distribution. A method is developed for estimating, from such a large sample with a side condition, all the missing parameters of the original distribution of $X_{(N)}$, that is the expectations, variances and covariances of $X_1, \cdots, X_p$, and the covariances $\sigma X_j X_k$ for $j = 1, \cdots, p$ and $k = p+1, \cdots, N$. This method does not require the knowledge of the side-condition. (This paper was prepared under the sponsorship of the Office of Naval Research.)

2. A Test of the Hypothesis that a Sample of Three Came from the Same Normal Distribution. Carl A. Bennett, General Electric Company, Hanford Works, Richland, Washington.

In the control of the precision of chemical analyses performed in duplicate, a test sometimes becomes necessary as to whether three determinations can reasonably be assumed to have arisen from the same normal population. A critical region for testing this hypothesis is given by $R > R_0$, where $R = D/d$, D being the maximum and d the minimum difference between the three values, and R_0 is determined by integration over the upper tail of the Cauchy distribution. It can easily be seen that this test is equivalent to a t-test between a sample of one and a sample of two.

3. A Note on the Application of the Abbreviated Doolittle Solution to Non-Orthogonal Analysis of Variance and Covariance. Carl A. Bennett, General Electric Company, Hanford Works, Richland, Washington.

S. S. Wilks has shown that the sums of squares necessary to the tests commonly made in non-orthogonal analyses of variance or covariance can in general be reduced to the ratio of two determinants. If several determinantal operations are performed to remove the singular principal minors, the abbreviated Doolittle solution yields these sums of squares directly. A combination of this technique and the calculational methods advanced by Wald and Yates greatly reduces the tedium of calculation in this type of analysis.

4. Yield Trials with Backcrossed Derived Lines of Wheat. G. A. Baker and F. N. Briggs, University of California, Davis.

Strains of White Federation 38 and Baart 38 Wheats derived by backcrossing sufficient to insure a high degree of homogeneity for all genetic factors were grown in conventional yield trials. The results were somewhat contradictory and led to a critical examination of such trials. The assumption that the deviations of yields in field trials from the specified pattern are random with uniform variance and expectation zero is not sufficiently realistic. We are led to consider a mathematical model which assumes a set of fertility levels upon

which a random element is superimposed. On the basis of this model it is possible to account for the low observed correlations between residuals and plot yields. In such a model the variance ratio F may be approximately unbiased but then its variance is smaller than under conventional assumptions. On the other hand, the expected value of F may be greater than one and sufficiently large so that "significant differences" between strains will always be found due to the differences in fertility levels. In such cases the results of the experiment may be misinterpreted. Transformations, in the ordinary sense of the word, will not bring such data into conformity with the conventional model. In order to bring the correlation between residuals and plot yields down to a sufficiently low level it is necessary to concentrate most of the variation in fertility levels into a few plots. That this is not unreasonable is borne out by agronomic observations. This model also explains the absence of correlation between the yields of strains as determined in two different trials on the same set of strains.

5. The Selection of the Largest of a Number of Means.

CHARLES M. STEIN, University of California, Berkeley.

Suppose X_{ij}, $i = 1, \cdots, p$; $j = 1, 2, \cdots$ are independently normally distributed with means $\xi_i + \eta_j$ and variances σ_j^2 where ξ_i, η_j are unknown but σ_j^2 are known. ϵ, α are fixed numbers with $0 < \epsilon, 0 < \alpha < 1$. It is desired to select, by a sequential procedure, in which we take first the observations with second subscript 1, etc. an integer M among $1, \cdots, p$ such that, for every $k = 1, \cdots, p$ and $\xi_1, \cdots \xi_p, \eta_1, \eta_2, \cdots$ satisfying $\xi_k \geqq \xi_j + \epsilon$ for all $j \neq k, P\{M = k\} \geqq 1 - \alpha$. In accordance with the following rule, one decides at each stage (after the observations with second subscript n) to take no more observations with certain first subscripts. For each $n = 1, 2, \cdots$ and each $l = 1, \cdots, p$ compute

$$\sum_{j=1}^{n} \frac{1}{\sigma_j^2}\left(X_{l,j} - \bar{X}_j - \frac{\epsilon(t_j - 1)}{t_j}\right)$$

where $\bar{X}_j$ is the average of the observations with second subscript j and t_j is the number of such observations. Continue taking observations $X_{l,n+1} \cdots$ for those l for which this expression is greater than $(ln\alpha)/\epsilon$ but not for the others. Eventually there will be at most one subscript $l = 1, \cdots, p$ for which one continues to take observations and if there is one this is chosen to be M. If there is none, the l for which the sum is largest is chosen to be M. This procedure is a straight-forward application of the Lemma on p. 146 of Wald's *Sequential Analysis*, and generalizations can easily be found.

6. The Effect of Inbreeding on Height at Withers in a Herd of Jersey Cattle.

W. C. ROLLINS, S. W. MEAD, and W. M. REGAN, University of California, Davis.

The data consist of measurements of height at withers of about 200 females for various ages from one month to five years. The intensity of inbreeding as measured by Wright's coefficient of inbreeding averaged 15 per cent and reached as high as 44 per cent in a few cases.

An intra-sire covariance analysis of height and per cent of inbreeding was made for various ages from the first month to the fifty-fourth month.

The results of the statistical analysis indicate that the inbred animals are shorter at one month of age and grow more slowly up to about the sixth month than do the outcrossed animals, but that from the sixth month on the inbreds begin to catch up with the outcrossed so that at maturity there is no significant difference in height.

7. An Example of a Singular Continuous Distribution. HENRY SCHEFFÉ, University of California at Los Angeles.

Simple and "natural" examples of singular continuous probability distributions are of pedagogical interest. They are trivially available in the k-variate case for $k > 1$. A univariate example may be obtained from the notion of a sequence of independent trials of an event with constant probability p of success, a notion familiar to the student and indispensable in elementary probability theory. The (real-valued) random variable X is taken to be the dyadic representation of the sequence of results (1 and 0, respectively, for success and failure). It is known that X has a singular continuous distribution for $p \neq 0, \frac{1}{2}, 1$. This result may be proved by using only the Tchebycheff inequality together with the formulas for the mean and variance of the binomial distribution.

8. On the Theory of Some Non-Parametric Hypotheses. ERICH L. LEHMANN and CHARLES STEIN, University of California, Berkeley, California.

For two types of non-parametric hypotheses optimum tests are derived against certain classes of alternatives. The two kinds of hypotheses are related and may be illustrated by the following example: (1) The joint distribution of the variables $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ is invariant under all permutations of the variables; (2) the variables are independently and identically distributed. It is shown that the theory of optimum tests for hypotheses of the first kind is the same as that of optimum similar tests for hypotheses of the second kind. Most powerful tests are obtained against arbitrary simple alternatives, and in a number of important cases most stringent tests are derived against certain composite alternatives. For the example (1), if the distributions are restricted to probability densities, Pitman's test based on $\bar{y} - \bar{x}$ is most powerful against the alternatives that the X's and Y's are independently normally distributed with common variance, and that $E(X_i) = \xi$, $E(Y_i) = \eta$ where $\eta > \xi$. If $\eta - \xi$ may be positive or negative the test based on $| \bar{y} - \bar{x} |$ is most stringent. The definitions are sufficiently general that the theory applies to both continuous and discrete problems, and that tied observations present no difficulties. It is shown that continuous and discrete problems may be combined. Pitman's test for example, when applied to certain discrete problems, coincides with Fisher's exact test, and when $m = n$ the test based on $| \bar{y} - \bar{x} |$ is most stringent for hypothesis (1) against a broad class of alternatives which includes both discrete and absolutely continuous distributions.

9. Concerning Compound Randomization in the Binary System. JOHN E. WALSH, Project Rand, Santa Monica, California.

Consider a set of binary digits. The numerical deviation from $\frac{1}{2}$ of the conditional probability that a specified digit equals 0 is called the bias of that digit for the given conditions on the remaining digits of the set. The maximum bias of the set is defined to be the maximum of the biases of the digits of the set. A set of binary digits is called random if its maximum bias is zero. Now consider an array of $(1 + t_1) \cdots (1 + t_K) \times n$ binary digits such that the rows are statistically independent. A compounding method of obtaining a set of $t_1 \cdots t_K n$ binary digits from the original array is presented. By suitable choices of $K, t_1, \cdots, t_K$ the maximum bias of the compounded set can be made extremely small even if the maximum bias of the original array is not small; this can be done so that $t_1 \cdots t_K / (1 + t_1) \cdots (1 + t_K)$ is moderately large. Also a method is outlined for constructing an approximately random binary digit table. This table has the property that the maximum bias of a set of digits taken from the table is an increasing function of the number of digits in the set.

10. A Multiple Decision Problem Arising in the Analysis of Variance. EDWARD PAULSON, University of Washington, Seattle.

In some applications of the analysis of variance, a procedure is required for classifying varieties into 'superior' and 'inferior' groups. Consider K varieties, with $x_{i\alpha}$ the α^{th} observation on the i^{th} variety ($\alpha = 1, 2, \cdots r; i = 1, 2, \cdots K$), let $\bar{x}_i = \sum_{\alpha=1}^{r} x_{i\alpha}/r$ and let s^2 be an independent estimate of the variance. For the i^{th} variety form the corresponding interval $\left(\bar{x}_i - \frac{\lambda s}{\sqrt{r}}, \bar{x}_i + \frac{\lambda s}{\sqrt{r}}\right)$. The superior group then consists of the variety with greatest sample mean, together with those varieties whose corresponding intervals have at least one point in common with the interval for the variety with the greatest mean. If all varieties fall into one group, this group is labeled 'neutral' and the varieties are considered homogeneous. To select λ, consider the relative importance of different incorrect classifications. For a given λ, an explicit expression is found for $P(A)$, the probability the varieties will not all be classified in one group when $m_1 = m_2 = \cdots = m_k$ where $m_i = E(\bar{x}_i)$; also explicit expressions are found for $P(B_1)$ and $P(B_2)$, where $P(B_1)$ is the probability there will not be a superior group consisting only of the Kth variety and $P(B_2)$ is the probability there will not be a superior group consisting of at least the Kth variety, when $m_1 = m_2 = \cdots = m_{k-1} = m$ and $m_k = m + \Delta(\Delta > 0)$. Similar results are obtained for classifying K processes according to their variances.

11. Recurrence Formulae for the Moments and Semi-variants of the Joint Distribution of the Sample Mean and Variance. OLAV REIERSØL, University of Oslo, Norway.

Let $x_1, x_2, \cdots, x_n$ be independent and having the same distribution. We consider the arithmetic mean m and the variance $v = (1/(n-1)) \Sigma (x_i - m)^2$. Let κ_{rs} denote the seminvariants of the joint distribution of m and v, and let the seminvariant generating operators K be defined by the equations: $\kappa_{r+1,s} = K_1\kappa_{rs}$, $\kappa_{r,s+1} = K_2\kappa_{rs}$ $K_i.1 = 0, K_i(PQ) = P(K_iQ) + Q(K_iP)$. An operator which operates only on the first factor of a product shall be denoted by a prime, and an operator which operates only on the second factor shall be denoted by a double prime. We have the following general formula, valid for any parent distribution: $K_1^r[(n-1)(K_2 + \kappa_{01}' + \kappa_{01}'') - 2n(K_1' + \kappa_{10}')(K_1'' + \kappa_{10}'')]^s[1 \cdot (\kappa_{01} - n\kappa_{20}) + n(\kappa_{10} \cdot \kappa_{10} - 1 \cdot \kappa_{10}^2)] = 0$. For $s = 0, 1, 2$, we obtain the formulae, $K_1^r(\kappa_{01} - n\kappa_{20}) = 0$, $K_1^r[(n-1)(\kappa_{02} - n\kappa_{21}) - 2n^2\kappa_{20}^2] = 0$, $K_1^r[(n-1)^2(\kappa_{03} - n\kappa_{21}) - 8n^2(n-1)\kappa_{21}\kappa_{20} + 4n^3(n-1)\kappa_{30}^2 - 8n^3(n-1)\kappa_{20}^3] = 0$.

12. The Problem of Identification in Factor Analysis. OLAV REIERSØL, University of Oslo, Norway.

The paper is concerned with the multiple factor analysis of L. L. Thurstone. Thurstone has given criteria which he says are almost certain to constitute sufficient and more than necessary conditions for uniqueness (i.e. identifiability) of a simple structure. It is shown that Thurstone's criteria are not always sufficient, and conditions are derived which are more nearly necessary and sufficient for the identifiability of a simple structure. Let A be the matrix of factor loadings with n rows and r columns. When the communalities are identifiable, the conditions will be: (1) Each column of A should have at least r zeros. (2) Let us consider the submatrix B of A, consisting of all the rows which have zeros in the kth column. Then, for $q = 1, 2, \cdots, r-1$, there should for any combination of q columns different from the kth, exist at least $q + 1$ rows of B containing non-zero elements in the q columns. This should be true for any value of k.

13. Note on Distinct Hypotheses. AGNES BERGER, Columbia University, New York.

As was pointed out by Neyman, one of the difficulties which one may encounter when devising a test to distinguish between two exhaustive and exclusive composite hypotheses referring to the unknown distribution of a random vector X is the following: If H_0 states that the true distribution function of X belongs to a set $\{F\}$ and H_1 that it belongs to a set $\{G\}$, it may happen that to every Borel set W of the sample space there exists an element F_W in $\{F\}$ and an element G_W in $\{G\}$ for which the probability of the sample point x falling on W is the same and therefore independent of whether H_0 or H_1 is true. If this is the case the pair H_0, H_1 is called non-distinct, otherwise they are called distinct. The existence of non-distinct hypotheses is demonstrated by a simple example, H_0 consisting of one, H_1 of three suitably chosen stepfunctions. It is shown however that if the sets $\{F\}$ and $\{G\}$ contain only continuous distribution functions and are at most enumerable then the pair H_0, H_1 is distinct. Necessary and sufficient conditions for H_0 and H_1 to be distinct were obtained jointly with Wald for an important class of hypotheses each containing a continuum of alternatives.

14. Place of Statistical Sampling in the Education of Engineers. E. L. GRANT, Stanford University.

There is convincing evidence that many engineering problems could be solved better with the aid of statistical methods than they are now solved without this aid. However, few practising engineers or teachers of engineering have had any training in statistical methods. As a result, those engineering problems which are in part statistical problems are seldom recognized as such. Even in the field of industrial quality control, in which successful applications of some of the simpler statistical techniques have been made in many different industries, the surface has barely been scratched and a serious obstacle to progress is the lack of a widespread appreciation of the statistics point of view among design engineers, production engineers, inspection personnel, and management.

This condition might gradually be corrected if during the next few years instruction in statistics should be introduced into all undergraduate engineering curricula. However, some recent discussions touching on the subject of statistics instruction for engineering students (e.g., the report on "The Teaching of Statistics" which appeared in the March 1948 issue of the *Annals of Math. Stat.*) have been most unrealistic regarding the amount of statistics instruction which could be added to engineering curricula. These discussions have suggested a full year of basic statistics followed by one or more courses in engineering applications. Desirable as this arrangement might be from the point of view of the most effective instruction in statistics, it is out of the question when considered in the light of the many subjects which are needed in engineering curricula. Although undergraduate engineering curricula have always been tighter than other curricula, the pressures today are greater than ever before—for more time devoted to the humanistic-social stem, for more time in basic mathematics and science, for introductory courses in various economic and management subjects such as engineering economy, accounting, industrial relations, business law, and industrial organization and management, and for more time in the various departmental courses in engineering subjects in order to permit presentation of important recent developments in engineering technology. Under these circumstances the most that can be hoped for in the undergraduate program is a single statistics course for one term, possibly three units for one semester or four units for one quarter. This should be supplemented by additional statistics instruction for some graduate students in engineering. A few engineering graduates should be encouraged to take graduate degrees in statistics and to make careers in the field of applied statistics.

In a successful undergraduate statistics course for engineering students, the problems and illustrations should be selected with two purposes in mind. One purpose, of course, should be to develop the principles of probability and statistical method. The other, equally important if these engineering graduates are to persuade their colleagues and superiors to adopt the statistics point of view in approaching engineering problems, should be to demonstrate how statistical method provides a useful guide to action in many different engineering situations. Applications of statistics to industrial quality control provide particularly good problems and illustrative examples which serve this second purpose.

15. Statistical Problems of Medical Diagnosis. JERZY NEYMAN, University of California, Berkeley.

"Diagnosis" is used to describe the outcome of a strictly defined test T, such as Wassermann test, which may lead to either of two possible outcomes, "positive" or "negative". Cases contemplated are such that at the time the test T is performed it is impossible to verify its verdict for certain and the best one can do is to repeat the test. It is postulated that to each individual of a population there corresponds a probability p that the test T will give a positive outcome. The value of p may vary from one individual to another. It is presumed that as p increases, the illness in the patient increases. Problem of comparison of two alternative tests and problem of estimating the distribution of p reduces to problems relating to the distribution of X = number of positive outcomes in n independent diagnoses. Statistical machinery suggested is that of *BAN* estimates (*Public Health Report*, Vol. 62, (1947), p. 1449). Principal result reported is that, with the mathematical model used in the paper quoted, the empirical variances of four *BAN* estimates computed for 205 samples of 1000 elements each agreed reasonably with the theoretical asymptotic values. Empirical distributions of three of these estimates did not show deviations from normality. That of the fourth was non-normal. It seems therefore that the asymptotic procedure of *BAN* estimate may be adequate for similar analyses.

16. Power of Certain Tests Relating to Medical Diagnosis. C. L. CHIANG and J. L. HODGES, JR., University of California, Berkeley.

Associate with each individual in a population π the probability p that he will be found tubercular when examined by a standard X-ray technique. Yerushalmy and others [*J. Am. Med. Assn.*, Vol. 133, (1947), p. 359] performed 5 independent such diagnoses on each of 1256 persons. Neyman [*Public Health Reports*, Vol. 62, (1947), p. 1449] proposed a simple four-parameter model for the distribution of p in π, estimated the parameters from the data of Yerushalmy and others, and obtained a satisfactory fit. In the present paper, the work of Neyman is paralleled with four new models, all giving satisfactory fit with the same data. The five models differ considerably in shape, and in the number of repeated diagnoses which they indicate to be necessary to detect a high proportion of those individuals having, say, $p \geqq 0.1$. Therefore further preliminary study seems indicated before one can design a mass survey to detect a high proportion of such persons. The approximate power of the χ^2 test of the Neyman model is considered, using one of the other models as alternative. It is found that to obtain power 0.7 with level of significance 0.05, it would be necessary to diagnose 5290 individuals 5 times each.

17. Iterative Treatment of Continuous Birth Processes. T. E. HARRIS, Project Rand, Santa Monica, California.

Random variables z_n are defined by $z_0 = 1$; $P(z_1 = r) = p_r$, $r = 1, 2, \cdots$; if $z_n = k$, z_{n+1} is the sum of k independent variates, each distributed like z_1. Let $x = \sum_1^\infty rp_r < \infty$;

$\sum_1^\infty r^2 p_r < \infty$; $0 < p_1 < 1$. The generating function $f(s) = \sum_1^\infty p_r s^r$ is said to be C.I. if there exists a family of generating functions $f(s, t)$ with $f(s, 1) = f(s)$, $f[f(s, t), t'] = f(s, tt')$ for all nonnegative t and t'. A necessary and sufficient condition that $f(s)$ be C.I. is that the numbers a_r, $r = 2, 3, \cdots$, be nonnegative; the a_r are determined recursively by requiring that the power series $\xi(s) = -s + \sum_2^\infty a_r s^r$ satisfy formally the functional equation $\xi(s)f'(s) = \xi[f(s)]$. The problem is connected with classical works on iteration. If $f(s)$ is C.I., the given Markoff process can be imbedded in a continuous birth process. If $\xi(s)$ is given, the m.g.f. $\phi(s)$ of the asymptotic distribution of the variate z_n/x^n may be determined from the formula $\phi^{-1}(s) = (s - 1) \exp \left\{ \int_1^s \left[\frac{\xi'(1)}{\xi(y)} + \frac{1}{1 - y} \right] dy \right\}$. Various properties of the corresponding distribution can be inferred from this expression.

18. Estimation of Means on the Basis of Preliminary Tests of Significance. Blair M. Bennett, University of California, Berkeley.

This paper examines the statistical procedure of pooling two sample means on the basis of the results of one or more preliminary tests of significance. Let x_i, $(i = 1, \cdots, N_1)$, represent a sample of N_1 observations from a normal population $\eta_1(\xi, \sigma_1^2)$, and y_i a sample of N_2 observations from $\eta_2(\eta, \sigma_2^2)$. An estimate of ξ which is commonly used in certain practical situations is given by: $x' = \bar{x}$, or $x' = (N_1\bar{x} + N_2\bar{y})/(N_1 + N_2)$, according as the sample means $\bar{x}$, $\bar{y}$ do or do not differ significantly on the basis of a preliminary test. The distribution of the estimate x' is determined, according as $\sigma_1 = \sigma_2$ are known or unknown. In both situations, the maximum (or minimum) bias is computed as a function of various levels of significance of the preliminary test of equality of means. Also, the mean square error of the estimate x' is calculated in both cases. If now equality of variances cannot be assumed, but an F-test of the sample variances s_1^2, s_2^2 does not indicate any significant difference, then in practice $\bar{x}$, $\bar{y}$ may be pooled, the weights being inversely proportional to the sample variances. Thus, the usual estimate of ξ will be of the form: $x' = \bar{x}$, or $x' = (N_1\bar{x}/s_1^2 + N_2\bar{y}/S_2^2)/(N_1/S_1^2 + N_2/S_2^2)$, according as $\bar{x}$ and $\bar{y}$ do or do not differ significantly on the basis of the Student t-test, subsequent to an F-test. The bias and mean square error of this estimate have been computed with the aid of the conditional power function of the t-test subsequent to an F-test.

19. Note on Power of the F Test. Stanley W. Nash, University of California, Berkeley.

Assuming "treatment" expectations to be normal random variables, the ratio of the sum of squares due to treatments to the sum of squares due to error has a central F distribution in the cases of randomized blocks, Latin squares, and one-way classifications. The F statistic converges in probability to a constant as the number of treatments is increased. This is one plus a multiple of the variance between treatment expectations. The power of the F test increases monotonely to one as the number of treatments is increased. This power can be calculated using tables of the incomplete beta function.

20. Best Asymptotically Normal Estimates. E. W. Barankin and J. Gurland, University of California, Berkeley.

The methods of minimum χ^2 developed by Neyman for obtaining BAN (best asymptotically normal) estimates of the parameters appearing in the multinomial distribution

are generalized to obtain certain optimum types of estimates in the case of an arbitrary distribution under certain restrictions. Let the random vector X have the probability density $v(x;\theta)$ in the absolutely continuous case and let $v(x;\theta) = P\{X = x/\theta\}$ in the discrete case, where θ is a fixed vector in the parameter space. Functions $\phi_i(X)$, $(i = 1, 2, \cdots, r)$ are selected for the purpose of forming estimates; these estimates are taken to be functions of the sample moments $\frac{1}{n}\sum_{j=1}^{n}\phi_i(x_j)$. Certain quadratic forms which depend on the choice of functions $\phi_1(X), \phi_2(X), \cdots, \phi_r(X)$ are minimized with respect to the parameters. In this manner, asymptotically normal estimates are obtained which are consistent, and have minimum asymptotic variances within the class of estimates so determined by the particular functions $\phi_1, \phi_2, \cdots, \phi_r$. It is possible, through a modification of this procedure, to obtain estimates by solving a set of linear equations. If $v(x;\theta)$ has the form

$$v(x;\theta) = \exp\{\beta_0(\theta) + \sum_{i=1}^{s}\beta_i(\theta)y_i(x) + y_0(x)\}$$

it can be shown that the best choice of the ϕ's is $y_1(x), y_2(x), \cdots, y_s(x)$. Maximum likelihood estimates belong to this class of BAN estimates.

BOOK REVIEW

The Theory of Games and Economic Behavior *John von Neumann and Oskar Morgenstern.* Princeton University Press, 1947; Second Edition, Pp. xviii, 641. $10.00

Reviewed by Leonid Hurwicz[1]

Iowa State College

This review is devoted to the second edition of a book which from its first appearance was acknowledged to be a major contribution in the field of theory of rational behavior. As is pointed out in the Preface, "the second edition differs from the first in some minor respects only". The main change is the addition of a proof (of "measurability" of utility) omitted in the first edition.

The book's objective is to solve the problem of rational behavior in a very general type of situation.

It is, therefore, not surprising that its results are of relevance in many fields of knowledge, among them economics and statistical inference.

In both economics and statistics the problem of rational behavior is a fundamental one. Thus one of the classical problems treated by the economic theory is that of profit maximization by a firm. The firm is assumed to be maximizing its net profit which is a function of prices of the product, materials used, etc., as well as the quantities used and produced. In the simplest case prices are taken as given; more generally they are assumed to be functions (known to the firm) of the quantities sold and purchased. But assuming this function to be known presupposes the knowledge of behavior of other firms. This procedure has for a long time been regarded as highly unsatisfactory; it is analogous to elaborating the theory of rational behavior of a poker player on the assumption that he knows the strategy of the other players!

It is the type of situation in which not only the behavior of various individuals, but even their strategies, are interdependent, that is treated by von Neumann and Morgenstern. The essence of their solutions is to base the optimal strategy on the *minimax principle*. As applied to a game, the principle requires that one should choose a strategy which minimizes the maximum loss that could be inflicted by the opponent.

The minimax principle, when applied by both players need not, in general, lead to a stable solution. To ensure the existence of such a solution the authors are led to the postulate that the choice of strategies be made through a random process. The minimax to be found is that of the *mathematical expectation* of the loss in the game. The latter postulate is of a restrictive nature[2] since it implies that the game is played for numerical ("measurable") stakes and that

[1] On leave to the United Nations Economic Commission for Europe.

[2] See Jacob C. Marschak, "Neumann's and Morgenstern's New Approach to Static Economics", *The Journal of Political Economy*, Vol. LIV (1946).

the second and higher moments of the probability distribution of the losses are immaterial. This restriction, however, has permitted the authors to go deeper in other directions. Given the great complexity of the problem, even in its restricted version, the authors' decision can hardly be criticized. One could only wish that similar considerations had made the authors more tolerant towards other work in the field of economics than is shown in some sections of the book.

The readers of the *Annals* will be particularly interested in the connection between the *Theory of Games* and the theory of statistical inference.

As has been pointed out by Abraham Wald[3] the problem faced by the statistician is somewhat similar to that of a player in a game of strategy. The theory of statistical inference may be viewed as a theory of rational behavior of the statistician. His "strategy" consists in adopting an optimal test or estimate, more generally an optimal decision function. This optimal decision function must be chosen without the knowledge of the "a priori" distribution of the population parameters. Wald's basic postulate of minimization of maximum risk is equivalent to regarding the statistician as a player in a game of strategy, with "Nature" as the other player. The optimal decision function is chosen in a way which (as shown by Wald) is equivalent to assuming the "least favorable" a priori distribution of the parameters. As Wald says, "we cannot say that Nature wants to maximize [the statistician's risk]. However, if the statistician is completely ignorant as to Nature's choice, it is perhaps not unreasonable to base the theory of a proper choice of [the decision function] on the assumption that Nature wants to maximize (the statistician's risk)".

It may be noted, however, that statistical inference, as seen by Wald, is a relatively simple game since it involves only two players and is of the zero-sum variety.

The admiring and enthusiastic reception given to the book's first edition would make any further general appraisal somewhat anticlimatic. Suffice it to say that a good deal of valuable work has already been stimulated by the *Theory of Games*, both in the field of social sciences and in mathematics.

[3]Abraham Wald, "Statistical Decision Functions which Minimize the Maximum Risk", *Annals of Mathematics*, Vol. 46, (1945).

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Dr. Paul H. Anderson, formerly an Economist with the War Assets Administration, Washington, D. C., has been appointed Professor of Marketing at Loyola University, New Orleans, Louisiana.

Mr. N. H. Carrier has resigned his position with the Mathematical Statistics Section, Chief Scientific Advisers Division, Ministry of Works, England to accept an appointment as Statistician in the General Register Office, Somerset House, Strand, London, W. C. 2, England.

Dr. T. Freeman Cope has been promoted to a full professorship at Queens College, Flushing, New York.

Dr. Wayne W. Gutzman, who was formerly at the Postgraduate School, Naval Academy, Annapolis as an Assistant Professor, has accepted a professorship in the Department of Mathematics, University of South Dakota.

Mr. Elvin A. Hoy has transferred from the position as Chief, Statistics Section, Bureau of Research and Statistics in the Social Security Administration to the position as Chief, Research Evaluation Section, Naval Reserve Training Publications, Navy Department, Naval Gun Factory, Washington, D. C.

Dr. Joe J. Livers has been promoted to a full professorship at Montana State College, Bozeman, Montana.

Professor Ernest S. Keeping has returned to his position at the University of Alberta, Edmonton, Alberta, Canada after having spent the spring term of 1948 at the Institute of North Carolina.

Mr. Wharton F. Keppler of the M&R Dietetic Laboratories, Inc., Columbus, Ohio has recently qualified as a Professional Industrial Engineer in the State of Ohio.

Mr. Ralph Mansfield has formed his own company to manufacture electrical testing equipment. The company is known as the Auto-Test, Incorporated, with Mr. Mansfield acting as Vice-president and Chief Engineer.

Mr. Jack Moshman has resigned an instructorship in mathematics at the University of Tennessee to accept the post of Statistician to the Medical Advisor, United States Atomic Energy Commission, Oak Ridge, Tennessee.

Mr. Bernard E. Phillips has resigned his position as teacher of mathematics in the Newark, New Jersey high schools to do statistical work for the Glenn L. Martin Co., Baltimore, Maryland.

Dr. W. R. Van Voorhis, Associate Professor of Mathematics, Fenn College, attended, as a representative of the Institute of Mathematical Statistics, the inauguration ceremonies of Dr. Keith Glennan as President of Case Institute of Technology, Cleveland, Ohio.

Atomic Energy Commission Fellowships

The National Research Council is announcing a new program of fellowships supported by funds provided by the Atomic Energy Commission as a part of the Commission's responsibility for future atomic energy research. Accordingly, these fellowships will be awarded in the many fields of science related to research in atomic energy.

A considerable number of these fellowships is available to young men and women who wish to continue in graduate training or research for the doctorate in an appropriate field of science. Others of these fellowships will provide training in biophysics applied to the control of radiation hazards. An additional number of fellowships will be assigned to those below the age of 35 who have already achieved the doctorate and who wish to secure advanced research training and experience in those aspects of the physical, biological and medical sciences related to atomic energy. Tenure of the fellowship does not impose on the fellow any commitment with regard to subsequent employment.

The candidates will be selected by the fellowship boards of the National Research Council established for this program. In the postdoctoral field, there will be three groups of fellowships, the basic stipend of which will be $3000. For the selection of fellows for advanced research and training in the general field of the physical sciences, a board has been established with Dr. Roger Adams, Professor of Chemistry, University of Illinois, as chairman. In the general field of the biological sciences, exclusive of the medical sciences, selections of postdoctoral fellows will be made by a board under the chairmanship of Dr. R. G. Gustavson, Chancellor of the University of Nebraska. For the selection of postdoctoral fellows in the medical sciences, a board has been set up under the chairmanship of Dr. Homer W. Smith, Professor of Physiology, College of Medicine, New York University.

The program provides for two groups of fellows in the predoctoral field, with stipends ranging from $1500–2100. One group of fellows will work in the biological and basic medical sciences including applied biophysics related to the measurement and control of radiation hazards and the effect of radiation upon health. Selections will be made by a board under the chairmanship of Dr. Douglas Whitaker, Professor of Biology, and Dean of the School of Biological Sciences, Stanford University. Another group of predoctoral fellows will be selected for study and research in the general field of the physical sciences. Selections will be made by a board under the chairmanship of Dr. Henry A. Barton, Director of the American Institute of Physics.

Fellowships will be granted for study and research in universities or other nonprofit research establishments approved by the fellowship boards. Awards will be made for the academic year 1948–49. Supervision of a fellow's program of work will be under the direction of the fellowship boards of the National Research Council. Further information can be secured by writing to the Fellowship Office, National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

Research Fellowships in Psychometrics

The Educational Testing Service, Princeton, N. J., is offering for 1949–50 its second series of research fellowships in psychometrics leading to the Ph.D. degree at Princeton University. Open to men who are acceptable to the Graduate School of the University, the two fellowships carry a stipend of $2,200 a year and are normally renewable.

Fellows will be engaged in part-time research in the general area of psychological measurement at the offices of the Educational Testing Service and will, in addition, carry a normal program of studies in the Graduate School. Competence in mathematics and psychology is a prerequisite for obtaining these fellowships. Information and application blanks may be obtained from: Director of Psychometric Fellowship Program, Educational Testing Service, Box 592, Princeton, N. J.

Preliminary Actuarial Examinations

Prize Awards

The winners of the prize awards offered by the Actuarial Society of America and the American Institute of Actuaries to the nine undergraduates ranking highest on the combined score on Part 1 and Part 2 of the 1948 Preliminary Actuarial Examinations are as follows:

First Prize of $200

Edward H. Larson.............................*Massachusetts Institute of Technology*

Additional Prizes of $100

John E. Brownlee.............................*Haverford College*
William L. Farmer.............................*University of Alabama*
Joseph P. Fennell.............................*Princeton University*
Bert F. Green, Jr..............................*Yale University*
Solomon Leader.............................*Rutgers University*
Felix A. E. Pirani.............................*University of Western Ontario*
Richard J. Semple.............................*University of Toronto*
Charles A. Yardley.............................*Dartmouth College*

The two actuarial organizations have authorized a similar set of nine prize awards for the 1949 Examinations on Part 2.

The Preliminary Actuarial Examinations consist of the following three examinations:

Part 1. *Language Aptitude Examination.*

(Reading comprehension, meaning of words and word relationships, antonyms, and verbal reasoning.)

Part 2. *General Mathematics Examination.*

(Algebra, trigonometry, coordinate geometry, differential and integral calculus.)

Part 3. *Special Mathematics Examination.*

(Finite differences, probability and statistics.)

The 1949 Preliminary actuarial Examinations will be administered by the Educational Testing Service at centers throughout the United States and Canada on May 13–14, 1949. The closing date for applications is March 15, 1949.

Detailed information concerning the Examinations can be obtained from either of the following organizations:

American Institute of Actuaries,
135 South LaSalle Street,
Chicago 3, Illinois.

The Actuarial Society of America,
393 Seventh Avenue,
New York 1, New York.

New Members

The following persons have been elected to membership in the Institute

(March 1 to May 31, 1948)

Alder, Arthur, Ph.D. (Univ. of Berne) Professor of Actuarial Science, University of Berne, Schlaeflistrasse 2, Berne, Switzerland.

Andrews, Fred C., B.S. (Univ. of Washington) Research Fellow, Department of Mathematics, University of Washington, *141 Savery Hall, University of Washington, Seattle, Washington.*

Archer, John, Actuary, Pensions Section, Lever Brothers and Unilever Ltd., *5A Spencer Hill, Wimbledon, S. W. 19, England.*

Benitz, Paul A., M.A. (Stanford Univ.) *173 Serpentine Road, Tenafly, New Jersey.*

Bennett, George K., Ph.D. (Yale) President of the Psychological Corporation, 522 Fifth Avenue, New York 18, New York.

Berrettoni, J. N., Ph.D. (Univ. of Minnesota) Professional Consultant in Statistics and Economics, *632 Erie St., S. E., Minneapolis 14, Minnesota.*

Birnbaum, Allan, A.B. (Univ. of Calif., Los Angeles) Teaching Assistant, Mathematical Statistics Department, Columbia University, *500 Riverside Drive, Room 434, New York 27, New York.*

Blank, Mark, M.A. (Univ. of Pennsylvania) Instructor of Philosophy, University of Pennsylvania, *223 E. Sedgwick, Philadelphia, Pa.*

Blumen, Isadore, B.A. (Univ. of Minn.) Student, Department of Mathematical Statistics, University of North Carolina, Chapel Hill, North Carolina.

Burdick, Wayne E., M.A. (Univ. of Mich.) Student, University of Michigan, *314 S. Fifth Avenue, Ann Arbor, Michigan.*

Chaturvedi, Jagdish C., M.Sc. (Agra Univ., India) Lecturer in Statistics, St. John's College, *37, Delhi Gate, Agra, U.P., India.*

Cote, Louis J., A.M. (Univ. of Mich.) Student, University of Michigan, *315 North State Street, Ann Arbor, Michigan.*

Dunleavy, Mary, A.B. (Hunter College, New York) Statistician, E. I. Dupont de Nemours, *657 Second Avenue, New York 16, New York.*

Ferber, Robert, M.A. (Univ. of Chicago) Student, University of Chicago, 54 West 89th Street, New York 24, New York.

Forman, John W., M.S. (Univ. of Iowa) Graduate Assistant, Department of Mathematics, State University of Iowa, Iowa City, Iowa.

Franklin, Nathan M., M.S. (Univ. of Mich.) Student, Univ. of Michigan, Box 195, *Moodus, Connecticut.*

Fraser, Donald A. S., M.A. (Univ. of Toronto) Instructor in Statistics, Graduate College, Princeton, New Jersey.

Grabowski, Edwin F., A.B. (George Washington Univ.) Student, George Washington University, *1330-30th Street, N.W., Washington, D. C.*

Healy, William C., Jr., B.S.E. (Univ. of Mich.) Student, University of Michigan, *589 Lincoln, Grosse Pointe, Michigan.*

Heimdahl, Olaf E. W., A.B. (Luther College, Washington) Teaching Fellow, Department of Mathematics, University of Washington, *4286 Union Bay Lane, Seattle 5, Washington.*

Henriksen, Robert O., B.Sc. (Univ. of Mich.) Student, University of Michigan, *751 Clancy Avenue, Grand Rapids, Michigan.*

Howard, William G., B.S. (Western Carolina Teachers College, Cullowhee, N. C.) Student, Institute of Statistics, University of North Carolina, *Route 1, Morrisville, North Carolina.*

Irick, Paul E., M.S. (Purdue Univ.) Mathematics Instructor, Purdue University, *729 North Grant St., West Lafayette, Indiana.*

Johnson, Elgy S., M.A. (Univ. of Mich.) Student, University of Michigan, *13907 Lincoln Street, Detroit 3, Michigan.*

Kaplan, E. L., B.S. (Carnegie Inst. of Tech.) Mathematician, Naval Ordnance Laboratory, *1427 N. St., N. W., Washington 5, D. C.*

Kaufman, Arthur, M.A. (Columbia Univ.) Student and Lecturer of Mathematics, Columbia University, *1280 Sheridan Avenue, New York 56, New York.*

Link, Richard F., B.S. (Univ. of Oregon) *750 W. Sixth St., Eugene, Oregon.*

Marks, Charles L., M.A. (Univ. of North Carolina) Instructor of Mathematics, University of North Carolina, *213 Mangum Dormitory, University of North Carolina, Chapel Hill, North Carolina.*

Marquardt, Mary, M.A. (Univ. of Illinois) Assistant Professor of Statistics, New York State School of Industrial and Labor Relations, Cornell University, Ithaca, New York.

Mickey, Max R., Jr., B.S. (Virginia Polytechnic Institute) Graduate Student and Graduate Assistant, Department of Mathematics, Iowa State College, *706 Ash Avenue, Ames, Iowa.*

Mindlin, Albert, B.A. (Univ. of California, Los Angeles) Teaching Assistant, Mathematics Department, University of California, *2444 Carlston Street, Berkeley 4, California.*

Morris, William S., A.B. (Princeton) Statistician, First Boston Corporation, 100 Broadway, New York 5, New York.

Netzorg, Morton J., Engineer, Development Tire Engineering Department, U. S. Rubber Co., Detroit, Michigan, *2523 Gladstone, Detroit 6, Michigan.*

Norton, James A., Jr., A.B. (Antioch College) Graduate Research Assistant, Veterans Guidance Center, Purdue University, West Lafayette, Indiana.

Perrin, John K., A.B. (Columbia College) Assistant Statistician, American Telephone & Telegraph Co., 195 Broadway, New York 7, New York.

Perry, Norman C., M.A. (Univ. of Southern Calif.) Lecturer in Mathematics, Mathematics Department, University of Southern California, Los Angeles, California.

Powell, Charles Jr., Actuary, Coates and Herfurth, Consulting Actuaries, 116 S. Virgil Avenue, Los Angeles 4, California.

Raiffa, Howard, B.S. (Univ. of Mich.) Student, University of Michigan, *1447 Enfield Court, Willow Run Village, Michigan.*

Raup, Joan E., B.A. (Barnard College) Statistical Analyst, Bureau of the Budget, *1436 N. Street, N. W., Washington 5, D. C.*

Rubinstein, David, B.S. (Univ. of Wash.) Research Assistant, Statistical Laboratory, University of California, *2216 Parker Street, Berkeley 4, California.*

Schlenz, John W., B.S. (Baldwin-Wallace College) Student, University of Michigan, *8306 Vineyard Avenue, Cleveland 5, Ohio.*

Scott, Elizabeth L., A.B. (Univ. of California) Research Assistant, Statistical Laboratory, Department of Mathematics, University of California, Berkeley 4, California.

Seidman, Herbert, B.A. (Brooklyn College) Junior Statistician, Chief, Statistical Information Section, New York University and Student, New York University, *2170 New York Avenue, Brooklyn 10, New York.*

Shaw, Oliver A., B.A. (Univ. of Mississippi) U.S. Air Force, *6431 Brooks Lane, N. W., Washington, D. C.*

Shellard, Gordon D., B.S. (Mass. Institute of Tech.) Assistant Section Head, Underwriting Studies Section, Actuarial Division, Metropolitan Life Insurance Co., *420 Mountain Avenue, Ridgewood, New Jersey.*

Shepherd, Clarence M., M.S. (Case Institute of Tech.) Electrochemical Research Chemist, *3959 Nichols Avenue, S. W., Washington, D. C.*

Shrikhande, Sharad-Chandra S., B.Sc. (Nagpur Univ., India) Graduate student, Department of Mathematical Statistics, University of North Carolina, Chapel Hill, North Carolina.

Sirlin, Robert, M.A. (Columbia Univ.) Statistician, Financial Analysis, *2046 East 23rd Street, Brooklyn 29, New York.*

Stacy, Edney W., A.B. (Univ. of North Carolina) Instructor of Mathematics, University of North Carolina, *301 W. Franklin Street, Chapel Hill, North Carolina.*

Sternhell, Charles M., B.S. (College of City of N. Y.) Section Head, Actuarial Division, Metropolitan Life Insurance Co., 1 Madison Avenue, New York City, New York.

Tang, Pei-Ching, Ph.D. (Univ. College, London Univ.) Professor, National Central University, Nanking, China.

Whitson, Milo E., A.M. (Geo. Peabody College, Nashville) Head of Mathematics Department, California State Polytechnic College, *523 Lawrence Dr., San Luis Obispo, California.*

Watson, Geoffrey S., B.A. (Univ. of Melbourne) Student, Institute of Statistics, State College, Raleigh, North Carolina.

Woolsey, Theodore D., B.A. (Yale Univ.) Biostatistician, Division of Public Health Methods, U. S. Public Health Service, *111 West Underwood St., Chevy Chase 15, Maryland.*

Wymer, John P., M.A. (Univ. of California, Berkeley) Statistician, U. S. Bureau of Labor Statistics, *719 Whittier St., N.W., Washington 12, D. C.*

Yerushalmy, Jacob, Ph.D. (Johns Hopkins Univ.) Professor of Biostatistics, School of Public Health, University of California, Berkeley 4, California.

REPORT ON THE BERKELEY MEETING OF THE INSTITUTE

The Thirty-fourth Meeting and the Third Regional West Coast Meeting of the Institute of Mathematical Statistics was held on the Berkeley Campus of the University of California June 22nd through June 24th, 1948, in conjunction with the Twenty-ninth Annual Meeting of the Pacific Division of the American Association for the Advancement of Science. During the meeting 115 persons registered, including the following members of the Institute:

G. A. Baker, Blair M. Bennett, Carl A. Bennett, Z. Wm. Birnbaum, David Blackwell, Albert H. Bowker, George W. Brown, A. George Carlton, Douglas G. Chapman, Andrew G. Clark, Edwin L. Crow, Dorothy Cruden, Harold Davis, R. C. Davis, W. J. Dixon, Robert Dorfman, George Eldredge, Lillian Elveback, Mary Elveback, Benjamin Epstein, M. W. Eudey, Evelyn Fix, Merrill M. Flood, H. H. Germond, Meyer A. Girshick, Eugene L. Grant, John Gurland, T. E. Harris, J. L. Hodges, Jr., Paul G. Hoel, John M. Howell, Harry M. Hughes, Leo Katz, H. S. Konijn, T. C. Koopmans, George W. Kuznets, E. L. Lehmann, Richard F. Link, A. M. Mood, Stanley W. Nash, J. Neyman, Stefan Peters, G. Baley Price, Kathryn B. Rolfe, Leonard J. Savage, Henry Scheffé, Howard L. Schug, Elizabeth L. Scott, Esther Seiden, Milton Sobel, Zenon Szatrowski, John W. Tukey, J. R. Vatnsdal, A. Wald, John E. Walsh, Holbrook Working, Zivia S. Wurtele.

The Tuesday morning session was devoted to a symposium on *Mathematical Theory of Games* with Professor G. C. Evans of the University of California, Berkeley, as chairman. Addresses were:

1. *Survey of von Neumann's mathematical theory of games.* J. C. C. McKinsey, Project Rand.
2. *Recent developments in the mathematical theory of games.* Olaf Helmer, Project Rand.
3. *Applications of theory of games to statistics.* Abraham Wald, Columbia University.
4. *On continuous games.* Henri F. Bohnenblust, California Institute of Technology.
5. *Discussion.* Edward W. Barankin, University of California, Berkeley.

The attendance was approximately 130.

The Tuesday afternoon session was under the chairmanship of Professor Henri F. Bohnenblust of the California Institute of Technology. The invited address, *Complete Classes of Statistical Decision Functions*, by Professor Abraham Wald was followed by tea in Senior Women's Hall and then the following contributed papers:

1. *Identification as a problem of inference.* T. C. Koopmans, Cowles Commission for Research in Economics.
   Discussion: Olav Reiersøl, University of Oslo.
2. *Estimation of parameters for truncated multinormal distributions.* Z. W. Birnbaum, E. Paulson and F. C. Andrews, University of Washington.
3. *A test of the hypothesis that a sample of three came from the same normal distribution.* Carl A. Bennett, General Electric Company.
4. *A Note on the application of the abbreviated Doolittle solution to nonorthogonal analysis of variance and covariance.* (By title.) Carl A. Bennett, General Electric Company.

The attendance was between 100 and 150 during the afternoon.

The Wednesday morning session was devoted to a symposium on *Design of Experiments with Particular Reference to Agricultural Trials.* Dean A. R. Davis of the University of California, Berkeley, presided briefly and then Professor Abraham Wald took over the duties of chairman. The papers were:

1. *Recent advances in experimental design.* R. C. Bose, University of Calcutta.
2. *Yield trials with backcrossed derived lines of wheat.* G. A. Baker and F. N. Briggs, University of California at Davis.
3. *Selecting subset which includes the largest of a number of means.* Charles Stein, University of California, Berkeley.
4. *Discussion.* A. G. Clark, Colorado State College; S. W. Nash, University of California, Berkeley; J. R. Vatnsdal, State College of Washington.
5. *The effect of inbreeding on height at withers in a herd of Jersey cattle.* W. C. Rollins, S. W. Mead and W. M. Regan, University of California at Davis.

Attendance was about 100.

The afternoon session, under the chairmanship of Professor George Pólya of Stanford University, began with an invited address by Professor Michel Loève, University of California, Berkeley, on *Random Functions and Related Problems.* This was followed by the contributed papers:

1. *An example of a singular continuous distribution.* (By title.) Henry Scheffé, University of California at Los Angeles.
2. *On the theory of some nonparametric hypotheses.* E. L. Lehmann and Charles Stein, University of California, Berkeley.
3. *Compound randomization in the binary system.* John E. Walsh, Project Rand.
4. *A multiple decision problem arising in the analysis of variance.* Edward Paulson, University of Washington.
5. *Recurrence formulae for the moments and seminvariants of the joint distribution of the sample mean and variance.* Olav Reiersøl, University of Oslo.
6. *Identification problem in factor analysis.* (By title.) Olav Reiersøl, University of Oslo.
7. *On distinct hypotheses.* Mrs. Agnes Berger, Columbia University.

The attendance was approximately 100.

A symposium on *Sampling for Industrial Use* occupied the Thursday morning session. Professor Z. W. Birnbaum of the University of Washington presided.

1. *Sampling plans for continuous production.* M. A. Girshick, Project Rand.
2. *Sampling plans with continuous variables for acceptance inspection.* A. L. Bowker, Stanford University.
3. *Place of statistical sampling in the education of engineers.* E. L. Grant, Stanford University.
4. *Discussion.* Henry Scheffé, University of California at Los Angeles; Charles Stein, University of California, Berkeley; Holbrook Working, Stanford University.

The attendance was approximately 100.

The first part of the afternoon session, presided over by Professor W. J. Dixon, University of Oregon, was devoted to contributed papers:

1. *Statistical problems of medical diagnosis.* Jerzy Neyman, University of California, Berkeley.
   Discussion: L. J. Savage, University of Chicago.

2. *Power of certain tests relating to medical diagnosis.* C. L. Chiang and J. L. Hodges, University of California, Berkeley.
3. *On best asymptotically normal estimates.* Edward W. Barankin and John Gurland, University of California, Berkeley.
4. *Iterative treatment of continuous birth processes.* T. E. Harris, Project Rand.
5. *Estimation of means on the basis of preliminary tests of significance.* Blair M. Bennett, University of California, Berkeley. (By title.)

The attendance was about 90.

The second part of the afternoon session was the Business Meeting. Professor Abraham Wald, President of the Institute, presided. It was recommended that two meetings a year be held on the West Coast, one in June in the San Francisco Bay Area alternating between Berkeley and Stanford and the other during the winter alternating between the North and Los Angeles. The next West Coast meeting will be held during the Thanksgiving recess at Seattle.